upwelling circulations in the ocean, yielding colder ocean surface temperatures and additional nutrients for biological growth.

The synergy of all these measurements makes the wind fields more valuable to science. Careful coordination of satellite research missions and ocean field programs (as in, for example, the World Ocean Circulation Experiment) were used in the past decade to maximize simultaneous data collection. These programs advanced our understanding of ocean circulation and atmosphere-ocean coupling while revealing the importance of temporal variability and trends. Given the difficulty of coordinating multiple satellite measurement programs and their vulnerability to system failure, even the best coordinated research missions are not a substitute for the sustained observing systems needed to understand and predict changes in climate and the role of ocean-atmosphere interactions in those changes. Considerable attention is now being focused on the design of sustained observing systems, including such programs as the Integrated Ocean Observing System and the National Polar-Orbiting Operational Environmental Satellite System (NPOESS).

Despite the phenomenal success of the SeaWinds scatterometer in measuring the small-scale wind patterns, the numerous advances made in understanding air-sea coupling with these observations, and a decade of experience with building, calibrating, and using the data for weather prediction, the future of this scatterometer is uncertain. Although the ESA-supplied advanced scatterometer (ASCAT) will be integrated into Europe's operational European Organisation for the Exploitation of Meteorological Satellites (EUMETSAT) MetOP missions, a scatterometer has not yet been included in any U.S.-planned monitoring system. Present NPOESS plans call for measurement of surface vector winds by polarimetric microwave radiometry (*8*), for which direction accuracy and all-weather measurement capability are not yet known. The low priority of wind directional accuracy is illustrated by NPOESS requirements: Wind speed is considered a key performance parameter (KPP), but wind direction is not. Further, no contingency plans exist for inclusion of scatterometers in NPOESS in the event that polarimetric radiometer measurements of vector winds prove inadequate for research or operational applications.

Wind and wind stress direction accuracy is critical to understanding the quantitative relationships between atmosphere and ocean and to operational requirements. For example, the upwelling shown in the figure as cold water near the coast is dynamically related to derivatives of the wind vectors, not to wind speed. For both atmospheric and ocean models, the dynamical boundary condition needed at the air-sea interface is stress, which requires the wind vectors. Wind stress magnitude is also not considered a KPP, and no directional accuracy requirement is specified for wind stress.

The European scatterometer measurements will be useful for many needs, despite lower spatial resolution and less coverage than SeaWinds. Nonetheless, proven, accurate, all-weather surface wind measurements need to be part of the U.S. operational program to improve weather and climate predictions. Given the long lead times required to plan and implement satellite observing systems, time is running out.

References

1. M. A. Donelan, W. J. Pierson, *J. Geophys. Res.* **92**, 4971 (1987).
2. D. B. Chelton, M. G. Schlax, M. H. Freilich, R. F. Milliff, *Science* **303**, 978 (2004); published online 15 January 2004 (10.1126/science.1091901).
3. D. Halpern, *J. Clim.* **1**, 1251 (1988).
4. M. H. Freilich, D. B. Chelton, *J. Phys. Oceanogr.* **16**, 741 (1986).
5. M. H. Freilich, R. S. Dunbar, *J. Geophys. Res.* **104**, 11231 (1999).
6. K. A. Kelly, S. Dickinson, M. J. McPhaden, G. C. Johnson, *Geophys. Res. Lett.* **28**, 2469 (2001).
7. K. A. Kelly, S. Dickinson, G. C. Johnson, in preparation.
8. S. H. Yueh, W. J. Wilson, S. J. Dinardo, F. K. Li, *IEEE Trans. Geosci. Remote Sensing* **37**, 949 (1999).

EVOLUTION

Transitions from Nonliving to Living Matter

Steen Rasmussen, Liaohai Chen, David Deamer, David C. Krakauer, Norman H. Packard, Peter F. Stadler, Mark A. Bedau

All life forms are composed of molecules that are not themselves alive. But in what ways do living and nonliving matter differ? How could a primitive life form arise from a collection of nonliving molecules? The transition from nonliving to living matter is usually raised in the context of the origin of life. Two recent international workshops (*1*) have taken a broader view and asked how simple life forms could be synthesized in the laboratory. The resulting artificial cells (sometimes called protocells) might be quite different from any extant or extinct form of life, perhaps orders of magnitude smaller than the smallest bacterium, and their synthesis need not recapitulate life's actual origins. A number of complementary studies have been steadily progressing toward the chemical construction of artificial cells (*2*–*8*). The two back-to-back workshops (*1*)—one held jointly at Los Alamos National Laboratory (LANL) and the Santa Fe Institute (SFI), and the other in Dortmund, Germany, at the seventh European Conference on Artificial Life—strived to encompass the full spectrum of this research.

There are two approaches to synthesizing artificial cells. The top-down approach aims to create them by simplifying and genetically reprogramming existing cells with simple genomes. Although the two workshops included a notable top-down exemplar, they focused primarily on the more general and more challenging bottom-up approach that aims to assemble artificial cells from scratch using nonliving organic and inorganic materials.

Although the definition of life is notoriously controversial, there is general agreement that a localized molecular assemblage should be considered alive if it continually regenerates itself, replicates itself, and is capable of evolving. Regeneration and replication involve transforming molecules and energy from the environment into cellular aggregations, and evolution requires heritable variation in cellular processes. The current consensus is that the simplest way to achieve these characteristics is to house informational polymers (such as DNA and RNA) and a metabolic system that chemically regulates and regenerates cellular components within a physical container (such as a lipid vesicle).

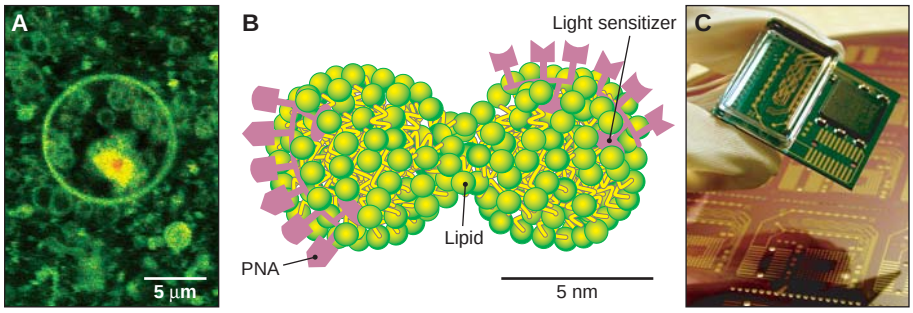
The workshops reviewed the state of the art in artificial cell research, much of which focuses on self-replicating lipid vesicles. David Deamer (Univ. of California, Santa Cruz) and Pier Luigi Luisi (ETH Zurich) each described the production of lipids using light energy, and the template-directed self-replication of RNA within a

S. Rasmussen is at Los Alamos National Laboratory, Los Alamos, NM 87545, USA. L. Chen is at Argonne National Laboratory, Argonne, IL 60439, USA. D. Deamer is at the University of California at Santa Cruz, Santa Cruz, CA 95064, USA. D. C. Krakauer is at the Santa Fe Institute, Santa Fe, NM 87506, USA. N. H. Packard is with ProtoLife Srl, Venice, Italy. P. F. Stadler is at the University of Leipzig, Leipzig, Germany. M. A. Bedau is at Reed College, Portland, OR 97202, USA. E-mail: steen@lanl.gov (S.R.)

lipid vesicle. In addition, Luisi demonstrated the polymerization of amino acids into proteins on the vesicle surface, which acts as a catalyst for the polymerization process. The principal hurdle remains the synthesis of efficient RNA replicases and related enzymes entirely within an artificial cell. Martin Hanczyc (Harvard Univ.) showed how the formation of lipid vesicles can be catalyzed by encapsulated clay particles with RNA adsorbed on their surfaces (see the figure) (*9*). This suggests that encapsulated clay could catalyze both the formation of lipid vesicles and the polymerization of RNA.

Successfully integrating different chemical systems is a key challenge in artificial cell research. Steen Rasmussen (LANL) and Liaohai Chen (Argonne National Laboratory) presented a minimal protocell design in which a small lipid aggregate (for example, a micelle) acts as a container by anchoring a lipophilic peptide nucleic acid (PNA, an analog of DNA with a pseudopeptide backbone) on its exterior (see the figure) (*7*). Peter Nielsen (Univ. of Copenhagen) amplified the benefits of using PNA as the informational polymer in such a system, and Kim Rasmussen (LANL) described experimentally and theoretically the hybridization instabilities and charge transfer in DNA-like double-helix systems. In the Chen-Rasmussen protocell, light-driven metabolic processes synthesize lipids and PNA, with the PNA acting as both an information molecule and as an electron-relay chain. This is the first explicit proposal that integrates genetics, metabolism, and containment in one chemical system. Metabolism in this system has been shown to produce lipids, but experimental realization of the rest of the integrated system has not yet been achieved.

The workshops also included experiments that use the top-down approach to artificial cell construction. Hamilton Smith (Institute for Biological Energy Alternatives, Rockville, Maryland) described an ongoing project to first simplify the genome of *Mycoplasma genitalium* (the organism with the simplest known genome), and then augment it with genes that encode proteins with desired functions. This top-down approach will probably produce the first impressive results because it can adopt wholesale the proven capacity of nature's biochemistry to integrate the complex reaction pathways crucial to cellular life, and because it relies on mature laboratory technology for DNA sequencing, synthesis, and manipulation. However, in the long run, the bottom-up approach might provide access to biochemical systems that are incompatible with or inaccessible using existing cellular chemistry, thus yielding a more diverse set of artificial cells with a wider range of useful properties.



When is a molecular aggregate alive? Examples of new experimental, computational, and technological advances in the development of artificial cells. (**A**) Short RNA oligonucleotides (red) are adsorbed to a particle of montmorillonite (clay) and encapsulated within a fatty acid vesicle (green). The assembly of RNA within the vesicle is coordinated by the clay particle (*9*). (**B**) A model of a dividing protocell, with an integrated metabolic, genetic, and container system (*7*). This minimal system consists of a carboxylic acid micelle with hydrophobic photosensitizer molecules and a lipophilic PNA, which doubles as a simple gene and as part of the metabolic system (electron relay). (**C**) The biomodule shown may be used for adaptive screening of flow conditions to evolve artificially compartmentalized cells (*10*). This module is based on microfluidic technology with online monitoring and fine-grained computer feedback control of fluxes (via electrodes in fluidic channels driven by reconfigurable logic). Such a computational biomodule interface could also be used to program the chemistry of artificial cells so that they could conduct useful tasks.

An emerging frontier in the development of artificial cells is the use of combinatorial chemistry to aid in the search for suitable chemical systems. John McCaskill (Fraunhofer-Gesellschaft, Germany) described technology that could integrate different chemical systems by developing chemical reactions across multiple spatially separated micrometer-sized channels, which act as computer-controlled microreactors. This technology could also provide "life-support" for artificial cells and their precursors, creating stepping stones toward autonomous artificial cells and enabling them to be programmed to perform useful functions. Statistics that enable open-ended evolution to be identified in data from evolving systems were described by Norman Packard (ProtoLife Srl) and Mark Bedau (Reed College, Portland, Oregon). Open-ended evolution is characterized by a continued ability to invent new properties—so far only the evolution of life on Earth (data partly from the fossil record) and human technology (data from patents) have been shown to generate adaptive novelty in an open-ended manner. Packard explained how statistics could be coupled with McCaskill's technology to automate the search for chemical systems that might be useful for artificial cells. Gunter von Kiedrowski (Ruhr-Univ. Bochum, Germany) described a new set of chemical reactions that use molecular elements he called "tetrabots." Such tetrabots could provide an important step toward replicating more general, spatially extended, DNA-like nanoarchitectures.

Several presentations described broader issues underpinning artificial cell theory, simulation, and experiment. Stirling Colgate (LANL), David Krakauer (SFI), Harold Morowitz (George Mason Univ., Virginia), and Eric Smith (SFI) attempted to clarify the distinction between nonliving and living matter. They described how nonliving chemical reactions, driven by thermodynamics, explore the state of space in an ergodical fashion, and thus tend to conduct a random exhaustive search of all possibilities; in contrast, living systems explore a combinatorially large space of possibilities through an evolutionary process. This echoed a central workshop theme: how and when information becomes a dominant factor in the evolution of life, that is, how and when selection plays a greater role than thermodynamics in the observed distribution of phenotypes. Peter Stadler (Univ. Leipzig) reviewed selection using replicator network dynamics, a theoretical framework describing population growth produced by different kinetic conditions. Smith and Morowitz further described how the citric acid cycle of living cells might be a thermodynamic attractor for all possible metabolic networks, thus explaining its appearance at the core of all living systems. Universal scaling in biological systems was discussed by Geoff West (SFI) and Woody Woodruff (LANL), who explained why regular patterns can be found, for example, between an organism's weight and metabolic rate, regardless of whether the organism is a bacterium or an elephant. Shelly Copley (Univ. of Colorado, Boulder) explained how catalysts operate in living systems today and how these were likely to have evolved from

CREDIT: (A) M. M. HANCZYC/MASSACHUSETTS GENERAL HOSPITAL; (B) S. RASMUSSEN/LANL AND SANTA FE INSTITUTE; (C) P. WAGLER *ET AL.*/FRAUNHOFER GESELLSCHAFT BIOMIP

less efficient precursors. Andrew Shreve (LANL) presented a rich variety of self-assembled nanomaterials that display specific emergent properties of a mechanical, photonic, or fluidic nature.

Computational methods are now powerful enough to suggest new experiments. Yi Jiang (LANL) reviewed the state of the art for molecular multiscale simulations in which the challenge is to connect realistic but slow molecular dynamic simulations with less accurate but fast higher level simulations. Andy Pohorille (NASA Ames Research Center, California) used simulations to argue that nongenomic early organisms could undergo evolution before the origin of organisms with genes. Takashi Ikegami (Univ. of Tokyo) presented simulations of a simple and abstract model of metabolic chemistry that demonstrates the spontaneous formation and reproduction of cell-like structures.

The workshops started with some tension between the origin of life perspective and the more general concern with synthesizing the simplest possible artificial cells. However, the participants eventually agreed that different artificial cell proposals might suggest different prebiotic niches. The workshop ended with a road-mapping exercise on four interrelated issues: (i) What is the boundary between physical and biological phenomena? (ii) What are key hurdles to integrating genes and energetics within a container? (iii) How can theory and simulation better inform artificial cell experiment? (iv) What are the most likely early technological applications of artificial cell research?

In time, research on these forms of artificial life will illuminate the perennial questions "What is life?" and "Where do we come from?" It will also eventually produce dramatic new technologies, such as self-repairing and self-replicating nanomachines. With metabolisms and genetics unlike those of existing organisms, such machines would literally form the basis of a living technology possessing powerful capabilities and raising important social and ethical implications. These issues were elaborated by Bedau, who suggested that the pursuit of these new technologies should be guided by what he called a "cautious courage" perspective. All workshop participants agreed that useful artificial cells will eventually be created, but there was no consensus about when.

References

1. www.ees.lanl.gov/protocells
2. C. Hutchinson *et al.*, *Science* **286**, 2165 (1999).
3. M. Bedau *et al.*, *Artif. Life* **6**, 363 (2000).
4. J. Szostak *et al.*, *Nature* **409**, 387 (2001).
5. A. Pohorille, D. Deamer, *Trends Biotechnol.* **20**, 123 (2002).
6. L. Eckardt *et al.*, *Nature* **420**, 286 (2002).
7 S. Rasmussen *et al.*, *Artif. Life* **9**, 269 (2003).
8. P. L. Luisi, *Origins Life Evol. Biosph.* **34**, 1 (2004).
9. M. M Hanczyc *et al.*, *Science* **302**, 618 (2003).
10. www.fraunhofer.de/english/press/pi/pi2003/03/rn_t7.html

PLANT SCIENCES

A *CONSTANS* Experience Brought to Light

John Klejnot and Chentao Lin

Each year during the spring, nature treats us to an amazing display of color and fragrance. Many plants bloom at this time of the year in response to seasonal changes of day length, a phenomenon called photoperiodism (*1*). Some plants, like Mendel's garden pea or today's experimental favorite *Arabidopsis*, flower more readily as days lengthen in the spring, whereas others such as rice or soybean prefer to flower when days get shorter in the fall. Since the discovery of photoperiodism in plants some 80 years ago (*1*), photoperiodic responses have been widely found in other organisms including mammals (*2*). How plants recognize photoperiods and respond to them by bringing forth blossoms has fascinated biologists for decades. On page 1003 of this issue, Valverde and colleagues (*3*) take us one step closer to understanding this phenomenon.

In plants, light signals are perceived by photoreceptors, which include phytochromes (phy) that respond to red/far-red light and cryptochromes (cry) that respond to blue/ultraviolet-A light (*4*). The light signals are "memorized" by the circadian clock and executed by transcription factors, which activate the floral meristem identity genes that initiate the transition from vegetative growth to reproductive development (*5*). Because neither the photoreceptors nor the circadian clock alone is sufficient to explain photoperiodic flowering, these components must somehow work together to measure day length changes (*5*). Almost a decade ago, Coupland's group discovered that an *Arabidopsis* gene called *CONSTANS* (*CO*) encodes a transcription factor that is critical for photoperiodic flowering (*6*). The CO protein activates the transcription of genes required for floral initiation, including a gene called *FLOWER LOCUS T* (*FT*) (*7*). The transcription of *CO* is governed by the circadian clock in a day length–dependent manner, and it has been hypothesized that a posttranscriptional regulatory mechanism must also be involved in regulating *CO* activity (*7*–*9*). Valverde *et al*. now show that the CO protein is ubiquitinated and then degraded by a protein complex called the proteasome, and that this process is regulated by both phytochromes and cryptochromes.

The researchers used transgenic *Arabidopsis* plants that constitutively express the *35S::CO* transgene independent of transcriptional control by the circadian clock and the *FT::LUC* reporter gene as a readout of *FT* promoter activity. They discovered that despite constitutive *CO* mRNA expression in the transgenic plants, the CO protein level was higher in the light phase of long days than in the light phase of short days, resulting in increased FT promoter activity during long days. Moreover, the abundance of CO protein and the activity of the *FT* promoter were greater in seedlings exposed to white, blue, or far-red light, relative to those exposed to red light or left in the dark. When recombinant CO protein was added to nuclear extracts of plant cells, it became ubiquitinated and degraded; degradation of CO was suppressed by proteasome inhibitors. Thus, CO is degraded in the dark via a ubiquitin/proteasome mechanism, and CO proteolysis is suppressed in light (blue and far-red).

To examine which photoreceptors are responsible for stabilizing CO in response to light, Valverde *et al.* crossed the *35S::CO* transgene into various photoreceptor mutant *Arabidopsis* plants. Analysis of the CO protein in the photoreceptor mutants showed that cry1/cry2 and phyA stabilize the CO protein in blue light and far-red light, respectively, and that phyB promotes CO degradation in red light. Apparently, these photoreceptors act to balance the abundance of CO protein in plants grown under conditions of natural light composed of different wavelengths (see the figure). Because both CO and FT activate flowering, these results also provide an explanation for why the *cry1cry2* and *phyA* mutants flower later than wild-type plants in blue light and far-red light, respectively; why the *phyB* mutant flowers earlier in red light; and how cry2 and phyA antagonize phyB action in white light to control flowering time (*10*, *11*). The

The authors are in the Department of Molecular, Cell and Developmental Biology, University of California, Los Angeles, CA 90095, USA. E-mail: clin@mcdb.ucla.edu

J|A|C|S COMMUNICATIONS

Published on Web 12/30/2008

Nucleobase Mediated, Photocatalytic Vesicle Formation from an Ester Precursor

Michael S. DeClue,[||,†] Pierre-Alain Monnard,[⊥,‡] James A. Bailey,[#,§] Sarah E. Maurer,[‡] Gavin E. Collis,[¶,†] Hans-Joachim Ziock,[‡] Steen Rasmussen,[*,⊥,‡] and James M. Boncella[*,†]

Material, Physics and Applications Division, Earth and Environmental Sciences Division, Chemistry Division, Los Alamos National Laboratory, P.O. Box 1663, Los Alamos, New Mexico 87545

Received October 17, 2008; E-mail: boncella@lanl.gov; steen@ifk.sdu.dk

It is well-known that amphiphiles will self-assemble into nanostructures under the right conditions. Notably, many fatty acids spontaneously form vesicles in water when the solution pH is adjusted close to the pK_a of the acid. This ability and their simple synthesis by Fischer−Tropsch reactions[1] under prebiotically plausible conditions have led to the suggestion that fatty acids preceded phospholipids as a primitive cellular (or cellular-like) membrane.[2−4]

Our interest is in the production of amphiphiles from precursor molecules and their subsequent self-assembly into vesicles because the dynamics of such a system may give insight into the types of chemical processes and couplings that could occur under prebiotic conditions. We have designed a chemical system in which fatty acid amphiphiles are generated from precursor molecules by visible light photolysis.

We report the use of photoinduced electron transfer to drive reductive cleavage of an ester to produce bilayer-forming molecules; specifically, visible photolysis in a mixture of a decanoic acid ester precursor, hydrogen donor molecules, and a ruthenium-based photocatalyst that employs a linked nucleobase (8-oxo-guanine) as an electron donor generates decanoic acid (Figure 1).[5,6] The overall transformation of the ester precursor to yield vesicles represents the use of an external energy source to convert nonstructure forming molecules into amphiphiles that spontaneously assemble into vesicles. The core of our chemical reaction system uses an 8-oxo-G-Ru photocatalyst, **2a**, a derivative of [tris(2,2′-bipyridine)-Ru(II)]$^{2+}$.

To understand the thermodynamics of this central catalytic reaction, one must consider the photoredox properties of [Ru(II)(bpy)$_3$]$^{2+}$, which are summarized in Figure 2 in the context of the chemistry employed in our system.[7] It is well-known that visible photoexcitation of [Ru(II)(bpy)$_3$]$^{2+}$ generates a metal-to-ligand charge transfer (MLCT) state that is both a better oxidant and reductant than the ground state.[7] This excited state has been used by many researchers to initiate electron transfer reactions.[8] Our system is designed to proceed via a reductive quenching pathway from the excited state (the left side of the diagram in Figure 2) and proceeds only if the donor possesses an oxidation potential less negative than ~ -1.0 V.

Guanine, the most easily oxidized conventional nucleobase, is not a sufficiently strong electron donor to provide an electron to the Ru MLCT excited state.[9] However, 8-oxo-G satisfies this requirement.[10] We attached (see Figure 1, compound **2a**) the 8-oxo-G to one of the bipyridine ligands via an electronically insulating alkane bridge to avoid significant changes in the redox potential from that of the parent [Ru(II)(bpy)$_3$]$^{2+}$ complex. An electron must be provided by the 8-oxo-G moiety (Figure 2, Reaction 1) generating the reactive element [Ru(II)(bpy^0)$_2$(bpy^{-1})]$^{1+}$ before the cleavage reaction can proceed. Once generated, the Ru^{1+} complex can then transfer charge to the picolinium ester providing carboxylic acid in the presence of a hydrogen source (Figure 2, Reaction 2).

Photolysis of a reaction mixture using visible light with all the components present gives substantial conversion to products within 24 h. The initial reaction mixture is a solution without visible structures (Figure 3A). As the photolysis proceeds, the solution becomes opalescent suggesting that large amphiphile structures are forming. Examination of these solutions via epifluorescence microscopy reveals that, at low conversions (<20 to ∼30%), oil-in-water emulsion compartments begin to form. By the time ∼44% conversion is achieved (Figure 3B), they are well established. When the critical vesicular concentration (CVC) is surpassed at ∼45−50% conversion, significant numbers of bilayer vesicular structures are observed along with oil-in-water emulsion compartments (Figure 3C). We observe that the CVC for the mixture of precursor ester and decanoic acid (DA) is considerably less than that of pure DA but have not observed the formation of vesicles with the precursor by itself under any conditions. At ≳65% conversions (Figure 3D) nearly all the structures that are observed are bilayer vesicles or tubules. These epifluorescence micrographs clearly show the evolution of the initially structureless system to one containing large numbers of membrane bilayer vesicles, demonstrating the spontaneous formation of organized structures from the precursor molecules. When the same experiment was performed using the guanine analogue of the photocatalyst (Figure 3E, 3F), no structures are observed with the exception of some of the fatty acid precursor that "phase-separated" from solution.

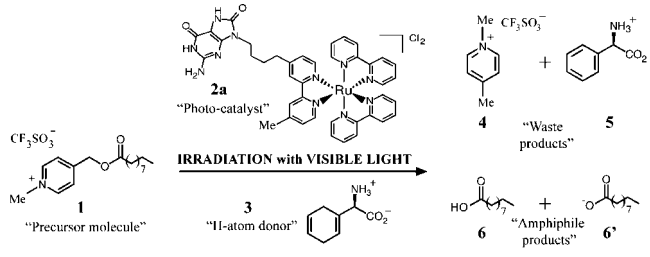


Figure 1. A summary of the nucleobase-mediated photochemical production of decanoic acid from its picolyl ester precursor.[5,6] The precursor **1** is dispersed in an aqueous buffered solution containing the electron donor linked photocatalyst **2a** and the dihydrophenylglycine hydrogen donor **3**. Upon irradiation, fatty acids [**6** protonated and **6′** deprotonated] form and self-assemble into membranous structures once the critical vesicle concentration of the precursor/fatty acid mixture is reached. Two "waste" compounds, *N*-methyl picolinium **4** and phenylglycine **5**, are also produced.

† Material, Physics and Applications Division, Los Alamos National Laboratory.
‡ Earth and Environmental Sciences Division, Los Alamos National Laboratory.
§ Chemistry Division, Los Alamos National Laboratory.
|| MDRNA Inc., Bothell, WA 98021.
⊥ Center for Fundamental Living Technology, Institute for Physics & Chemistry, University of Southern Denmark, Odense, Denmark.
Irving K Barber School of Arts and Sciences, UBC Okanagan, Kelowna, Canada.
¶ CSIRO Molecular and Health Technologies, Clayton, Australia.

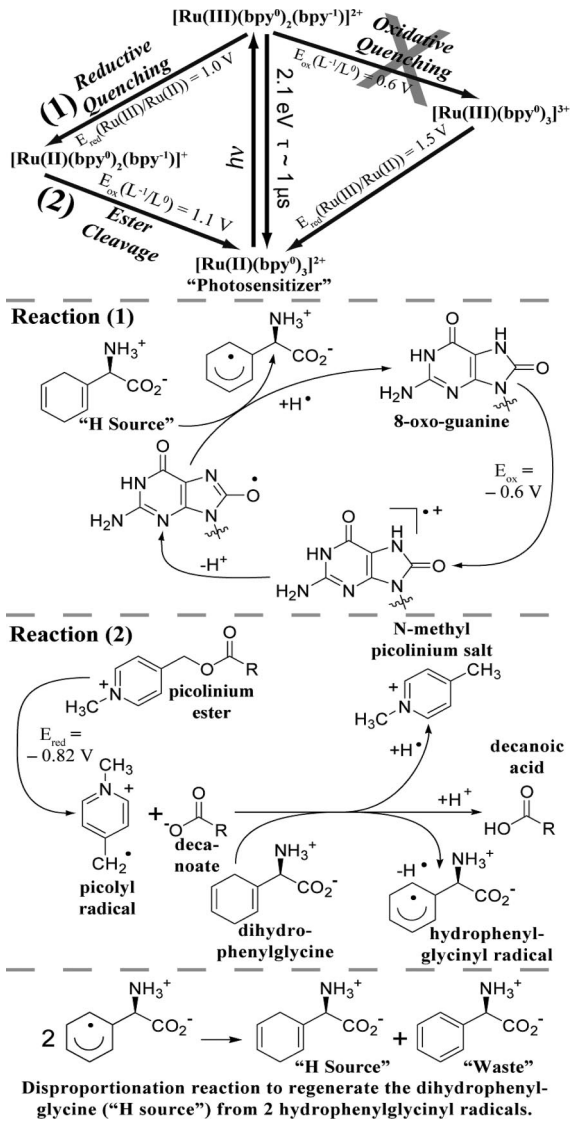


Figure 2. A summary of the reductive and oxidative pathways for the $[Ru(II)(bpy)_3]^{2+}$ complex showing the redox potentials and reactions of the photocatalyzed ester cleavage reaction. Only the reductive quenching pathway is thermodynamically feasible. All electrode potentials are relative to NHE from Vlcek et al.[5−7,10]

When the reaction was monitored using ^{1}H NMR spectroscopy (Figure 4), the clean conversion of a 15 mM concentration of precursor ester to the final products of the reaction was seen when using only a 1 mM concentration of the 8-oxo-G-Ru catalyst. Also, the NMR signals at ∼0.85, 1.25, and 1.60 ppm arising from the H atoms on the alkyl chains of the decanoate groups that are produced show significant broadening at ∼40% conversion, which becomes more pronounced as the reaction continues. At high conversions, these NMR signals are nearly broadened into the baseline as is expected if the alkyl chains of the amphiphile products have restricted motion due to their incorporation into organized structures.[11] These structures are destroyed when the reaction mixture is diluted with water (D_2O) such that the concentration of decanoic acid falls below its CVC (trace f, Figure 4).

We also performed the photocatalytic reaction using a 405-nm laser and monitored the reaction with FTIR spectroscopy. This system is more amenable to quantitative analysis of the changing concentrations of the reactants than the NMR experiments and has allowed us to collect reproducible kinetic data for the reaction. For the FTIR experiment, a short-path length cell (<30 μm) was used

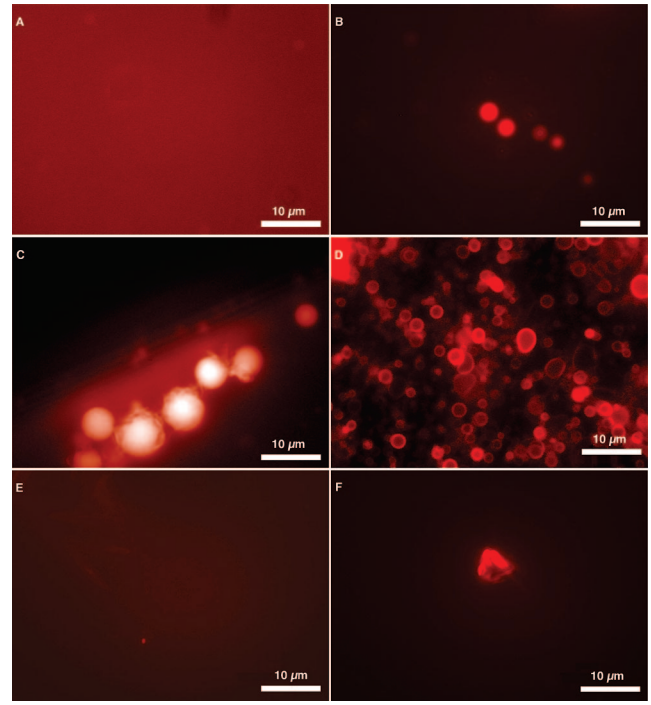


Figure 3. Epifluorescence micrographs of aggregate formation mediated by the photolysis reaction. All reaction samples are stained with 2.5 μM Nile Red, which strongly associates with hydrophobic molecular regions and permits the visualization of bilayer structures (D). The epifluorescence micrographs were taken under UV-illumination and observed through a 550−560 nm filter. (A−D) The 8-oxo-G linked photocatalyst: (A) at t = 0 h with 0% conversion; (B) after 6-h irradiation with ∼44% conversion, oil−water emulsion structures likely coated by fatty acid are freely moving in the medium; (C) after 8-h irradiation with ∼47% conversion, both emulsion and membranous structures are observed. Note the apparent shedding of material from the emulsion structures; (D) after ∼65% conversion at 24 h. Vesicles and other membranous structures are already readily apparent at ∼50% conversion (not shown). (E and F) The guanine analogue of the photocatalyst: (E) at t = 0 h with 0% conversion; (F) at t = 24 h. The bright spot visible is some of the fatty acid precursor that has "phase-separated" from solution.

because the resulting limited attenuation of the excitation beam yields an essentially flat excitation profile. This change results in an increased reaction rate relative to that of the NMR experiments, giving 50% conversion within 3 h. The results from two experiments using this short-path length cell with the decanoic acid precursor are shown in Figure 5, and the results from several other experiments are found in the Supporting Information (SI).

Several conclusions can be drawn from the kinetic data. Most importantly, when **2a** is used, the reaction proceeds smoothly, cleanly, and catalytically. When we attempted to use the analogous guanine linked complex (G-Ru) as the photosensitizer, only a small amount of decanoic acid was produced. The generation of *N*-methyl picolinium carbinol (5−7% conversion) as indicated by NMR results when the guanine complex was used as a photosensitizer suggests that the slow background hydrolysis of the ester becomes the predominant reaction in the guanine case. These results demonstrate that **2a** is essential for the reaction to occur and that the G-Ru species shows no catalytic rate enhancement, essentially duplicating the results obtained from the Ru sensitizer with no nucleobase in the system. Furthermore, the reaction is clearly dependent upon the presence of the Ru complex and a specific nucleobase (8-oxo-G).

The initial quantum yield for the reaction was determined to be 0.44% in the laser photolysis experiments of Figure 5 (full details in SI). Given that the reaction slows with time, this quantum yield

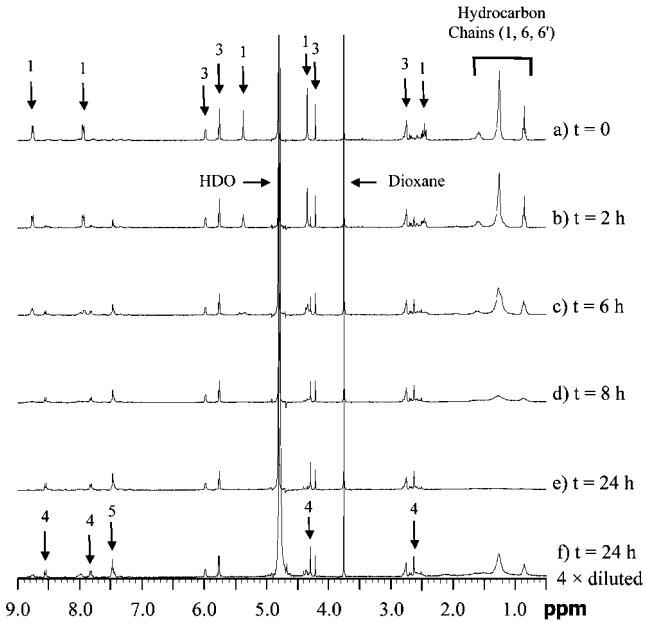


Figure 4. Photolysis of a 15.0 mM solution of ester precursor, with 1.0 mM of the nucleobase−catalyst complex, and 15.75 mM dihydrophenylglycine in D_2O (100 mM phosphate buffer at pD = 7.0) when illuminated with a 150 W tungsten spotlight, followed by ^1H NMR for different durations of exposure to light. The samples were photolyzed in a water bath maintained at 20 °C. All traces are normalized to the same dioxane peak integral value. For different times, the conversions are (a) 0%, (b) ∼22%, (c) ∼44%, (d) ∼47%, and (e) ∼65%. The bottom trace (f) shows the ^1H NMR of the same material as in (e), but after dilution and with the signal strength scaled to regain the initial dioxane calibration signal strength. The dilution destroys the vesicular structures allowing the alkyl signals from amphiphile products to return to the spectrum. Compound numbers refer to Figure 1.

is an upper limit. Other experiments using HPLC analysis of the product mixtures show that >90% of the Ru catalyst remains after 24 h of reaction. We have not explicitly measured the luminescence quantum yield of the catalyst or the guanine derivative but have observed that the luminescence lifetimes of these complexes are the same as that of the $[Ru(II)(bpy)_2(4,4'Me_2bpy)]^{2+}$ model complex at room temperature in solution. This suggests that quenching by the oxo-G moiety is inefficient and is consistent with the observed low quantum yield.

In summary, we have devised and demonstrated a chemical process in which an 8-oxo-G molecule directly mediates the photochemical conversion of an ester precursor into a fatty acid product that self-assembles into a vesicle. This is in contrast to earlier work in which a simple direct photolysis of a precursor resulted in vesicles.[12] The overall production of fatty acid molecules leading to the spontaneous formation of vesicles presented here is driven by an external energy source rather than by supplying activated components that spontaneously react to yield vesicle building blocks. The net result is a chemical reaction system that is controllable and creates structures capable of growth.

In order for this chemical system to exhibit more of the necessary functions of a self-replicating network, several further steps will clearly be necessary.[13] All the species must be colocated into/onto the container, the 8-oxo-G molecule must be incorporated into an oligomer or polymer with a base sequence that has the potential for templated replication[14] and kinetic regulation of the photolysis

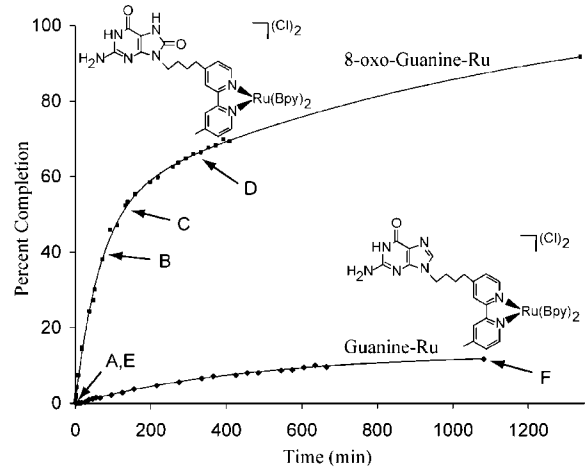


Figure 5. Kinetics of the reaction of decanoate precursor in the presence of 8-oxo-G-Ru catalyst (■) or G-Ru complex (◆). Conversion is calculated using the loss of the absorbance due to the ester carbonyl stretching vibration at 1740 cm^{-1}, an area free of interfering absorbance changes. There is no reaction beyond background with the guanine Ru complex. Labels A−F correspond to the same conversion levels shown in the images of Figure 3 but occur at different times due to changes in the experimental setup (light intensities, wavelengths, and pathlengths).

chemistry, and the system must continue to function under these conditions. Our ongoing research efforts are striving to address these issues.

Acknowledgment. We thank the Los Alamos National Laboratory LDRD program for financial support. We are also indebted to Dr. Liaohai Chen, Dr. William Woodruff, and Prof. Peter Nielsen for valuable technical suggestions and critical questions.

Supporting Information Available: A description of the synthetic procedures and physical data for the new compounds including NMR spectra is provided along with descriptions of the FTIR analyses used to monitor reaction progress. This material is available free of charge via the Internet at http://pubs.acs.org.

References

(1) McCollom, T. M.; Ritter, G.; Simoneit, B. R. T. *Orig. Life Evol. Biosphere* **1999**, *29*, 153–166.
(2) Hargreaves, W. R.; Deamer. D. W. Origin and early evolution of bilayer membranes. In *Light Transducing Membranes: Structure, Function and Evolution*; Deamer, D. W., Ed.; Academic Press: New York, 1978; pp 23−59.
(3) Luisi, P. L.; Walde, P.; Oberholzer, T. *Curr. Opin. Colloid Interface Sci.* **1999**, *4*, 33–39.
(4) Monnard, P.-A.; Deamer, D. W. *Anat. Rec.* **2002**, *268*, 196–207.
(5) Sundararajan, C.; Falvey, D. E. *Photochem. Photobiol. Sci.* **2006**, *5*, 116–121.
(6) Sundararajan, C.; Falvey, D. E. *J. Org. Chem.* **2004**, *69*, 5547–5554.
(7) Vlcek, A. A.; Dodsworth, E. S.; Pietro, W. J.; Lever, A. B. P. *Inorg. Chem.* **1995**, *34*, 1906–1913.
(8) Gray, H. B.; Winkler, J. R. *Proc. Natl. Acad. Sci. U.S.A.* **2005**, *102*, 3534–3539.
(9) Burrows, C. J.; Muller, J. G. *Chem. Rev.* **1998**, *98*, 1109–1151.
(10) Goyal, R. N.; Dryhurst, G. *J. Electroanal. Chem.* **1982**, *135*, 75–91.
(11) Sánchez de Groot, N.; Parella, T.; Aviles, F. X.; Vendrell, J.; Ventura, S. *Biophys. J.* **2007**, *92*, 1732–1741.
(12) Veronese, A.; Berclaz, N.; Luisi, P. L. *J. Phys. Chem. B* **1998**, *102*, 7078–7080.
(13) Rasmussen, S.; Chen, L. H.; Nilsson, M.; Abe, S. *Artificial Life* **2003**, *9*, 269–316.
(14) Lipscomb, L. A.; Peek, M. E.; Morningstar, S. M.; Verghis, S. M.; Miller, E. M.; Rich, A.; Essigmann, J. M.; Williams, L. D. *Proc. Natl. Acad. Sci. U.S.A.* **1995**, *92*, 719–723.

JA808200N

PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY B

Phil. Trans. R. Soc. B (2007) **362**, 1841–1845
doi:10.1098/rstb.2007.2076
Published online 1 May 2007

Emergence of protocellular growth laws

Tristan Rocheleau[1,2], Steen Rasmussen[1,3,4,*], Peter E. Nielsen[4], Martin N. Jacobi[5] and Hans Ziock[1]

[1]*Self-Organizing Systems, EES-6, Los Alamos National Laboratory, MS-D462, Los Alamos, NM 87545, USA*
[2]*University of California Santa Barbara, Santa Barbara, CA 93106, USA*
[3]*Santa Fe Institute, Santa Fe, NM 87501, USA*
[4]*University of Copenhagen, IMBG, Blegdamsvej 3, 2200 Copenhagen, Denmark*
[5]*Chalmers University of Technology, 41296 Gothenburg, Sweden*

Template-directed replication is known to obey a parabolic growth law due to product inhibition (Sievers & Von Kiedrowski 1994 *Nature* **369**, 221; Lee *et al*. 1996 *Nature* **382**, 525; Varga & Szathmáry 1997 *Bull. Math. Biol*. **59**, 1145). We investigate a template-directed replication with a coupled template catalysed lipid aggregate production as a model of a minimal protocell and show analytically that the autocatalytic template–container feedback ensures balanced exponential replication kinetics; both the genes and the container grow exponentially with the same exponent. The parabolic gene replication does not limit the protocellular growth, and a detailed stoichiometric control of the individual protocell components is not necessary to ensure a balanced gene–container growth as conjectured by various authors (Gánti 2004 *Chemoton theory*). Our analysis also suggests that the exponential growth of most modern biological systems emerges from the inherent spatial quality of the container replication process as we show analytically how the internal gene and metabolic kinetics determine the cell population's generation time and not the growth law (Burdett & Kirkwood 1983 *J. Theor. Biol*. **103**, 11–20; Novak *et al*. 1998 *Biophys. Chem*. **72**, 185–200; Tyson *et al*. 2003 *Curr. Opin. Cell Biol*. **15**, 221–231). Previous extensive replication reaction kinetic studies have mainly focused on template replication and have not included a coupling to metabolic container dynamics (Stadler *et al*. 2000 *Bull. Math. Biol*. **62**, 1061–1086; Stadler & Stadler 2003 *Adv. Comp. Syst*. **6**, 47). The reported results extend these investigations. Finally, the coordinated exponential gene–container growth law stemming from catalysis is an encouraging circumstance for the many experimental groups currently engaged in assembling self-replicating minimal artificial cells (Szostak 2001 *et al. Nature* **409**, 387–390; Pohorille & Deamer 2002 *Trends Biotech*. **20** 123–128; Rasmussen *et al*. 2004 *Science* **303**, 963–965; Szathmáry 2005 *Nature* **433**, 469–470; Luisi *et al*. 2006 *Naturwissenschaften* **93**, 1–13).[1]

Keywords: protocell integration; replication; metabolism; kinetics

1. LIPID-ASSOCIATED TEMPLATE-DIRECTED REPLICATION

Our simple protocell model consists of a lipid container (e.g. a large micelle) with amphiphilic replicating protogene molecules and an associated hydrophobic metabolic system as discussed in Rasmussen *et al*. (2003). However, the details of the coupling between the genes and the container are not important as long as the container defines a localized and restrictive volume or area for the genes (Szostak *et al*. 2001; Pohorille & Deamer 2002; Rasmussen *et al*. 2004; Szathmáry 2005; Luisi *et al*. 2006). In our current minimal representation, only the replicator and container kinetics are modelled explicitly. We can assume that all lipophilic replicators are bound to the exterior surfaces of the lipid aggregates and that both single- and double-stranded templates coexist at these lipid interfaces (figure 1*a*). For the following arguments, it is only necessary to localize the genes. In our model, we can, for example, assume that the replicator molecules are peptide nucleic acid (Pschl *et al*. 1998; Nielsen 1999; Mattes & Seitz 2001) oligomers with decorated (lipophilic) backbones. We can further assume that the single- and double-stranded templates are at equilibrium with each other as long as the template replication rate is much slower than the hybridization and dehybridization processes.

The local replication reaction can then be written as

$$T_s + \alpha O \rightarrow T_d \rightleftharpoons T_s + T_s, \tag{1.1}$$

where O is a resource replicator substrate molecule and α is the number of such molecules used in each replication. Denoting these local template concentrations by T_s^l and T_d^l for the single- and double-stranded templates, respectively, the local equilibrium condition for the single- and double-stranded templates is then given by

$$K_t = \frac{(T_s^l)^2}{T_d^l}, \tag{1.2}$$

where K_t defines the equilibrium constant. In the following, we assume that $K_t \ll 1$. If T_{tot}^l defines the total local template concentration, single- or double

* Author for correspondence (steen@lanl.gov).

One contribution of 13 to a Theme Issue 'Towards the artificial cell'.

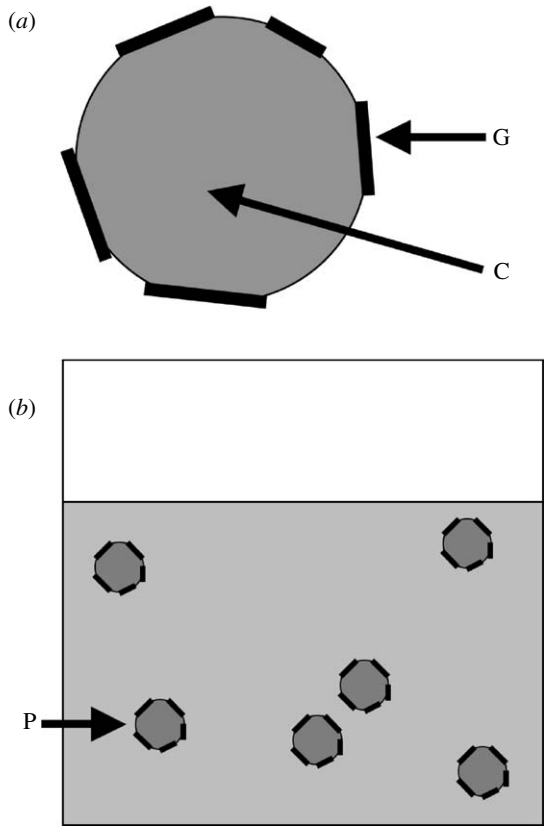


Figure 1. (*a*) A protocell consists of a lipid container C (e.g. a large micelle) with amphiphilic gene molecules G at local aggregate concentration $T^{\mathrm{l}}_{\mathrm{tot}}$ attached to the surface (Rasmussen *et al.* 2004). (*b*) The average aggregate of each protocell P is composed of m_0 lipid molecules. The (global) aggregate concentration is A and the global template concentration $T^{\mathrm{g}}_{\mathrm{tot}} = T^{\mathrm{l}}_{\mathrm{tot}} A V_{\mathrm{A}}$, where $T^{\mathrm{l}}_{\mathrm{tot}}$ denotes the local aggregate template concentration (template molecules per aggregate volume).

stranded, we have

$$T^{\mathrm{l}}_{\mathrm{tot}} = T^{\mathrm{l}}_{\mathrm{s}} + 2T^{\mathrm{l}}_{\mathrm{d}}. \tag{1.3}$$

$T^{\mathrm{l}}_{\mathrm{s}}$ and $T^{\mathrm{l}}_{\mathrm{d}}$ can now be expressed as a function of $T^{\mathrm{l}}_{\mathrm{tot}}$ by combining equations (1.2) and (1.3),

$$\begin{aligned} T^{\mathrm{l}}_{\mathrm{s}} &= \frac{-K_t + \sqrt{K_t^2 + 8K_t T^{\mathrm{l}}_{\mathrm{tot}}}}{4} \\ &= \sqrt{\frac{K_t T^{\mathrm{l}}_{\mathrm{tot}}}{2}} + O_1(K_t), \end{aligned} \tag{1.4}$$

$$\begin{aligned} T^{\mathrm{l}}_{\mathrm{d}} &= \frac{T^{\mathrm{l}}_{\mathrm{tot}}}{2} - \frac{T^{\mathrm{l}}_{\mathrm{s}}}{2} \\ &= \frac{T^{\mathrm{l}}_{\mathrm{tot}}}{2} - \frac{-K_t + \sqrt{K_t^2 + 8K_t T^{\mathrm{l}}_{\mathrm{tot}}}}{8} \\ &= \frac{T^{\mathrm{l}}_{\mathrm{tot}}}{2} - \sqrt{\frac{K_t T^{\mathrm{l}}_{\mathrm{tot}}}{8}} - \frac{O_2(K_t)}{2}, \end{aligned} \tag{1.5}$$

where we have used that $K_t \ll T_{\mathrm{tot}}$ ($K_t \ll 1$) and $O_i(K_t)$, $i = 1, 2$ is small in terms of K_t.

The left side of equation (1.1) defines the template-directed replication of $T^{\mathrm{l}}_{\mathrm{s}}$ using some background precursor template molecules with the local aggregate concentration O^{l}. The template replication rate equation then becomes

$$\frac{\mathrm{d}T^{\mathrm{l}}_{\mathrm{tot}}}{\mathrm{d}t} = k_{\mathrm{T}} f(O^{\mathrm{l}}) T^{\mathrm{l}}_{\mathrm{s}} = \frac{1}{\alpha} k_{\mathrm{T}} O^{\mathrm{l}} T^{\mathrm{l}}_{\mathrm{s}}, \tag{1.6}$$

where $f(\cdot)$ is a reaction mechanism dependent, which is O^{l}/α in the case where α resource oligomers are added sequentially to the template one at the time, while k_{T} is the single oligomer addition rate constant. Substituting equation (1.4) into equation (1.6)

$$\frac{\mathrm{d}T^{\mathrm{l}}_{\mathrm{tot}}}{\mathrm{d}t} = k_{\mathrm{T}} f(O^{\mathrm{l}}) \left(\sqrt{\frac{K_t T^{\mathrm{l}}_{\mathrm{tot}}}{2}} + O_1(K_t) \right), \tag{1.7}$$

where $O_1(K_t)$ is small, which to the leading order results in parabolic growth

$$\begin{aligned} T^{\mathrm{l}}_{\mathrm{tot}}(t) &= \frac{f(O^{\mathrm{l}})^2 K_t k_{\mathrm{T}}^2}{8} t^2 + \frac{\sqrt{T^{\mathrm{l}}_{\mathrm{tot}}(0) K_t} f(O^{\mathrm{l}}) k_{\mathrm{T}}}{\sqrt{2}} t + T^{\mathrm{l}}_{\mathrm{tot}}(0) \\ &= \left(\frac{f(O^{\mathrm{l}}) \sqrt{K_t} k_{\mathrm{T}}}{2\sqrt{2}} t + \sqrt{T^{\mathrm{l}}_{\mathrm{tot}}(0)} \right)^2, \end{aligned} \tag{1.8}$$

where $T^{\mathrm{l}}_{\mathrm{tot}}(0)$ is the concentration at $t = 0$. This result is an example of the well-known growth law for product inhibition in template-directed replication (Sievers & Von Kiedrowski 1994) in an unlimited system.

2. TEMPLATE CATALYSED CONTAINER REPRODUCTION

In our model, the aggregate-associated lipid production L is template catalysed and can be described by

$$pL + T_{\mathrm{d}} \rightarrow T_{\mathrm{d}} + L, \tag{2.1}$$

where pL is the precursor lipid. We assume that the catalysis is dominated by the double-stranded template (Rasmussen *et al.* 2003). The lipid production rate equation is then

$$\frac{\mathrm{d}L^{\mathrm{l}}}{\mathrm{d}t} = k_{\mathrm{L}} p L^{\mathrm{l}} T^{\mathrm{l}}_{\mathrm{d}}, \tag{2.2}$$

where k_{L} is the lipid production rate constant and pL^{l} is the local (background) lipid precursor concentration.

This simple gene–container coupling can be generalized to explicitly account for a variety of details involved in the metabolic processes as discussed in Munteanu *et al.* (2007). In this paper, we focus on a detailed discussion of the above minimal gene–container connection.

We assume the average number of lipid molecules per average aggregate volume V_{A} to be m_0. An average aggregate number can, for example, be maintained through a natural micellar size distribution (Rosen 1988). The relative local (dimensionless) aggregate growth rate can now be expressed as

$$\frac{V_{\mathrm{A}}}{m_0} \frac{\mathrm{d}L^{\mathrm{l}}}{\mathrm{d}t} = k_{\mathrm{L}} p L^{\mathrm{l}} T^{\mathrm{l}}_{\mathrm{d}} \frac{V_{\mathrm{A}}}{m_0}. \tag{2.3}$$

Let A be the global aggregate concentration. Since each aggregate is assumed to divide following growth in order to maintain an average of m_0 lipid molecules per aggregate, the growth in the global aggregate concentration due to the local lipid production in the aggregates can be expressed as

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \frac{A V_{\mathrm{A}}}{m_0} \frac{\mathrm{d}L^{\mathrm{l}}}{\mathrm{d}t} = k_{\mathrm{L}} p L^{\mathrm{l}} T^{\mathrm{l}}_{\mathrm{d}} \frac{A V_{\mathrm{A}}}{m_0}. \tag{2.4}$$

In the above expression, we have implicitly assumed that the volume of the templates and its change are small (constant) compared with the volume of lipid in

an aggregate. We can now transform the local aggregate concentration of templates into the corresponding global concentrations using the simple relation between local and global template concentrations

$$T^{\mathrm{g}} = T^{\mathrm{l}} A V_{\mathrm{A}}, \tag{2.5}$$

noting that T^{l} is measured in template molecules per lipid aggregate volume V_{A} and A is measured in the number of lipid aggregates (of average size V_{A}) per system volume. Substituting equation (2.5) into equation (2.4) and taking into account that $T_{\mathrm{tot}} \approx 2T_{\mathrm{d}}$, we obtain

$$\frac{\mathrm{d}A}{\mathrm{d}t} = \frac{k_{\mathrm{L}} p L^{\mathrm{l}}}{m_0} T^{\mathrm{g}}_{\mathrm{d}} = \gamma_{\mathrm{A}} T^{\mathrm{g}}_{\mathrm{tot}}, \tag{2.6}$$

where

$$\gamma_{\mathrm{A}} = \frac{k_{\mathrm{L}} p L^{\mathrm{l}}}{2m_0}. \tag{2.7}$$

It should be noted that we can just as well express the local concentrations in molecules per aggregate, in molecules per aggregate volume using an average aggregate volume of V_{A} or in molecules per aggregate surface area S_{A}.

3. GLOBAL TEMPLATE REPRODUCTION

The growth equation for the local total template concentration $T^{\mathrm{l}}_{\mathrm{tot}}$ corresponding to equation (1.6) is given by

$$\frac{\mathrm{d}T^{\mathrm{l}}_{\mathrm{tot}}}{\mathrm{d}t} = k_{\mathrm{T}} f(O^{\mathrm{l}}) T^{\mathrm{l}}_{\mathrm{s}} - \frac{\mathrm{d}A}{\mathrm{d}t} \frac{T^{\mathrm{l}}_{\mathrm{tot}}}{A}, \tag{3.1}$$

where the second term stems from the effect the volume change has on the local concentration. Since equation (3.1) is central to the argument, we discuss it in detail. The question that needs to be answered is what happens to the concentrations when one goes from a well-stirred solution to many small containers. This has to do with the origin of the second term in equation (3.1). The volume of the container grows with time and hence the concentration can change, and decreases with increasing container volume as A is measured in aggregates of average size V_{A} per litre. The volume growth of the container must be equal to the growth in the (global) concentration of the aggregates. One also needs a normalization factor to go from the change in container volume to the change in template concentration. All together, this yields a minus sign (the concentration goes down with increasing volume) multiplied by the ratio of the concentration of templates divided by the concentration of aggregates.

Now, since equation (1.2) holds for the local concentration (not the global), we can use the approximate form of equation (1.4) to rewrite the growth equation as

$$\frac{\mathrm{d}T^{\mathrm{l}}_{\mathrm{tot}}}{\mathrm{d}t} = k_{\mathrm{T}} f(O^{\mathrm{l}}) \sqrt{\frac{K_t T^{\mathrm{l}}_{\mathrm{tot}}}{2}} - \frac{\mathrm{d}A}{\mathrm{d}t} \frac{T^{\mathrm{l}}_{\mathrm{tot}}}{A}. \tag{3.2}$$

To rewrite the growth in terms of global concentrations, we multiply the growth equation (3.2) by A, and get

$$\frac{\mathrm{d}\left(AT^{\mathrm{l}}_{\mathrm{tot}}\right)}{\mathrm{d}t} - \frac{\mathrm{d}A}{\mathrm{d}t} T^{\mathrm{l}}_{\mathrm{tot}} = k_{\mathrm{T}} f(O^{\mathrm{l}}) \sqrt{\frac{K_t T^{\mathrm{l}}_{\mathrm{tot}}}{2}} A - \frac{\mathrm{d}A}{\mathrm{d}t} T^{\mathrm{l}}_{\mathrm{tot}}. \tag{3.3}$$

Note that the volume compensation term vanishes as it should. We then use the relation $T^{g} = T^{\mathrm{l}} A V_{\mathrm{A}}$ to rewrite the growth in terms of global concentrations

$$\frac{\mathrm{d}T^{\mathrm{g}}_{\mathrm{tot}}}{\mathrm{d}t} = \gamma_{\mathrm{T}} \sqrt{T^{\mathrm{g}}_{\mathrm{tot}} A}, \tag{3.4}$$

where

$$\gamma_{\mathrm{T}} = k_{\mathrm{T}} f(O^{\mathrm{l}}) \sqrt{\frac{K_t V_{\mathrm{A}}}{2}}. \tag{3.5}$$

4. COUPLED TEMPLATE AND CONTAINER REPRODUCTION

The coupled growth equations (2.6) and (3.4) can be solved by separation of variables, which can be used to rewrite equation (3.4) where we are seeking the solution for large $T^{\mathrm{g}}_{\mathrm{tot}}$ values. We obtain (Appendix A) the solution,

$$T^{\mathrm{g}}_{\mathrm{tot}}(t) = \frac{\left(\exp\left(\frac{3}{2}at + D_0\right) - b\right)^{2/3}}{a^{2/3}} \qquad \sim \exp(at), \tag{4.1}$$

where

$$a = \gamma_{\mathrm{T}}^{2/3} \gamma_{\mathrm{A}}^{1/3}, \tag{4.2}$$

and

$$b = \frac{\gamma_{\mathrm{T}}^{5/3} C_0}{3\gamma_{\mathrm{A}}^{2/3}}, \quad C_0 = A(0)^{3/2} - \frac{\gamma_{\mathrm{A}}}{\gamma_{\mathrm{T}}} T^{\mathrm{g}}_{\mathrm{tot}}(0)^{3/2}, \tag{4.3}$$

$$\text{and} \quad D_0 = \ln\left(aT^{\mathrm{g}}_{\mathrm{tot}}(0)^{3/2} + b\right).$$

Substituting equation (4.1) into equation (2.6) yields

$$\frac{\mathrm{d}A}{\mathrm{d}t} \sim \gamma_{\mathrm{A}} \exp(at), \tag{4.4}$$

which means that

$$A(t) \sim \exp(at) \sim T^{\mathrm{g}}_{\mathrm{tot}}(t). \tag{4.5}$$

Thus, to a leading order, both the coupled template and the aggregate grow exponentially with the same exponent (a). As equations (2.7) and (3.5) are substituted into equation (4.2), we note that the exponent in a direct manner depends on the template and lipid production rate constants, k_{T} and k_{L}, the local background precursor concentration, pL, the background template substrate concentration and the order of the template reaction, $f(O^{\mathrm{l}})$, the template duplex equilibrium constant, K_t and the average number of lipids per aggregate, m_0.

$$a = \left(\frac{1}{4} k_{\mathrm{T}}^2 k_{\mathrm{L}} p L^{\mathrm{l}} f(O^{\mathrm{l}})^2 \frac{K_t V_{\mathrm{A}}}{m_0}\right)^{1/3}. \tag{4.6}$$

In figure 2, a simulation of the original kinetic equations (2.6) and (3.4) illustrates these findings. It should be noted that at least one gene duplex must be present in each aggregate for the derived kinetic growth equations to be valid, which is ensured if $\gamma_{\mathrm{A}} < \gamma_{\mathrm{T}}$.

5. DISCUSSION

A kinetic analysis of a minimal protocell model shows that the local parabolic template growth law is compensated by parallel aggregate division dynamics, which results in an overall exponential growth for the

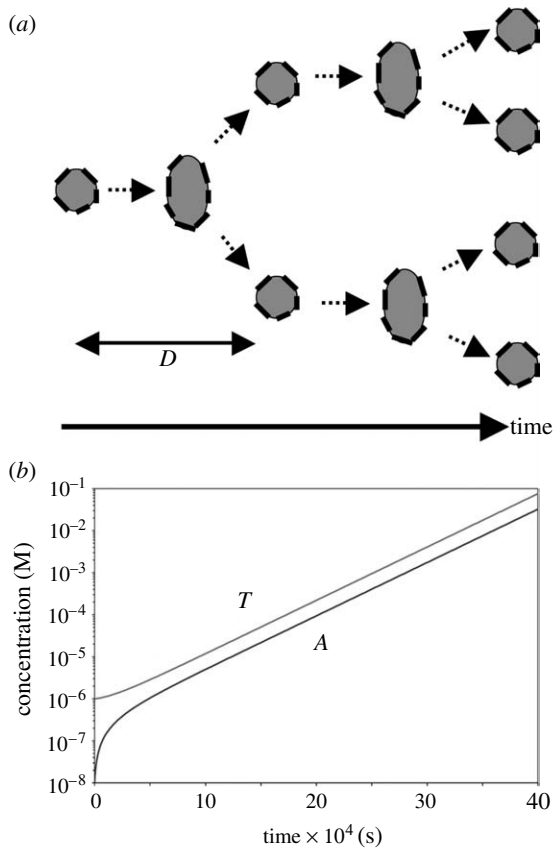


Figure 2. (*a*) The template catalyses the lipid production, while the lipid aggregate makes possible (catalyses) the template replication. Since the local template concentration is kept approximately constant due to the aggregate growth, the template replicates exponentially. The generation (doubling) time is given by $D = \left[(1/4)k_{\mathrm{T}}^2 k_{\mathrm{L}} f(O^{\mathrm{l}})^2 p L^{\mathrm{l}} (K_t V_A)/m_0\right]^{-(1/3)} \ln 2$. See §5 in the text. (*b*) The results of a numerical simulation of the original coupled template–container replication system equations (2.6) and (3.4) illustrating the derived coupled asymptotic exponential growth law. Parameter values: $m_0 \sim 1000$, $K_t V_{\mathrm{A}} \sim 10^{-3}$, $pL^{\mathrm{l}} \sim 10^{-3}$ M, $f(O^{\mathrm{l}}) \sim 2 \times 10^{-4}$ M, $k_{\mathrm{L}} \sim 25\ \mathrm{Ms}^{-1}$ and $k_{\mathrm{T}} \sim 10\ \mathrm{Ms}^{-1}$. Note how the gene and the container concentrations converge to a coordinated growth rate after an initial transient adjustment period, due to an unbalanced set of initial concentrations. The protocell population doubles approximately once a day with these parameter values.

whole system. In simple terms, the coupled replication/container growth through equations (1.1) and (2.1) assures that at steady-state conditions the average local value of the template concentration $T_{\mathrm{d}}^{\mathrm{l}}$ is kept constant as the volume of the system (the number of containers) grow concertedly. Consequently, no shift in the equilibrium of equation (1.1) for the system as a whole is occurring during growth and replication (in contrast to solution replication at constant volume). A second consequence is that the average concentration of single-stranded templates $T_{\mathrm{s}}^{\mathrm{l}}$ and thus the replication rate is kept constant, allowing the system as such (genes + container) to grow exponentially. This result holds for any coupled genetic growth law as well as for an explicit inclusion of the metabolic kinetics or any other detailed intracellular growth dynamics (Munteanu *et al.* 2007). We have expressed analytically how the generation time is affected by the internal gene replication kinetics.

Thus, a detailed stoichiometric growth control of the individual protocell components is not necessary as conjectured, for example, by Gánti (2004). A coupled autocatalytic feedback between the template and the aggregate replication processes generates a coordinated growth of the main components. This result is also an extension of the exponential growth law in the low template concentration limit as discussed by Stadler (Stadler & Stadler 2003) and others. As long as the effective [2] average local replicator concentration is kept constant, in this case by growing the system's total lipid volume, exponential growth is also possible. A discussion of the detailed growth dynamics in modern cells can be found in e.g. Burdett & Kirkwood (1983); Novak *et al.* (1998) and Tyson *et al.* (2003).

We thank the Protocell Assembly team (http://protocells.lanl.gov/team), Camille S.-O. Attolini, Jerzy Maselko and Andreea Munteanu for their critical feedback on the calculations and earlier versions of this manuscript. We also thank Eörs Szathmáry for pointing us to early work on intracellular growth kinetics. This work is supported in part by the Los Alamos National Laboratory LDRD-DR grant on 'Protocell Assembly' (PAs) and by the European Commission's Sixth Framework project on 'Programmable Artificial Cell Evolution' (PACE).

ENDNOTES

[1]It should be noted that the main result presented in this paper can be generalized as demonstrated in Munteanu *et al.* (2007).

[2]It should be noted that an exponential amplification is also possible in a naked template-directed replication process if the resulting new template is modified in the ligation process to lower its hybridization energy (Paul & Joyce 2004).

APPENDIX A

The coupled growth equations (2.6) and (3.4) can be solved by separation of variables, i.e. by dividing equation (3.4) by equation (2.6)

$$\frac{\mathrm{d}T_{\mathrm{tot}}^{\mathrm{g}}}{\mathrm{d}A} = \frac{\gamma_{\mathrm{T}}}{\gamma_{\mathrm{A}}} \frac{\sqrt{A}}{\sqrt{T_{\mathrm{tot}}^{\mathrm{g}}}}. \tag{A 1}$$

Note that equation (A 1) is a special case of a Bernoulli differential equation. From equation (A 1), we obtain

$$(A(t))^{3/2} = \frac{\gamma_{\mathrm{A}}}{\gamma_{\mathrm{T}}} \left(T_{\mathrm{tot}}^{\mathrm{g}}(t)\right)^{3/2} + C_0, \tag{A 2}$$

where $C_0 = A(0)^{3/2} - \dfrac{\gamma_{\mathrm{A}}}{\gamma_{\mathrm{T}}} T_{\mathrm{tot}}^{\mathrm{g}}(0)^{3/2}$. We use this to rewrite equation (3.4) as

$$\begin{aligned}\frac{\mathrm{d}T_{\mathrm{tot}}^{\mathrm{g}}}{\mathrm{d}t} &= \gamma_{\mathrm{T}} \sqrt{T_{\mathrm{tot}}^{\mathrm{g}}} \left(\frac{\gamma_{\mathrm{A}}}{\gamma_{\mathrm{T}}} \left(T_{\mathrm{tot}}^{\mathrm{g}}\right)^{3/2} + C_0\right)^{1/3} \\ &= \gamma_{\mathrm{T}} \sqrt{T_{\mathrm{tot}}^{\mathrm{g}}} \left(\frac{\gamma_{\mathrm{A}}^{1/3}}{\gamma_{\mathrm{T}}^{1/3}} \sqrt{T_{\mathrm{tot}}^{\mathrm{g}}} + \frac{C_0 \gamma_{\mathrm{T}}^{2/3}}{3\gamma_{\mathrm{A}}^{2/3} T_{\mathrm{tot}}^{\mathrm{g}}} + \mathrm{O}\left[\left(T_{\mathrm{tot}}^{\mathrm{g}}\right)^{-5/2}\right]\right),\end{aligned} \tag{A 3}$$

where we have expanded the expression for large $T_{\mathrm{tot}}^{\mathrm{g}}$ relative to C_0. Ignoring the last term, which is small in order $T_{\mathrm{tot}}^{\mathrm{g}-5/2}$, the simplified expression has the solution

$$T_{\mathrm{tot}}^{\mathrm{g}}(t) = \frac{\left(\exp\left(\frac{3}{2}at + D_0\right) - b\right)^{2/3}}{a^{2/3}} \sim \exp(at), \tag{A 4}$$

to the leading order in T^{g}_{tot}, where $a = \gamma_A^{1/3}\gamma_T^{2/3}$, $b = (\gamma_T^{5/3}C_0)/3\gamma_A^{2/3}$ and $D_0 = \ln(b + aT^{g}_{tot}(0)^{3/2})$. Substituting equation (4.1) into equation (2.6) yields

$$\frac{dA}{dt} \sim \gamma_A \exp(at), \tag{A 5}$$

which means that

$$A(t) \sim \exp(at) \sim T^{g}_{tot}(t). \tag{A 6}$$

Thus, to leading order, the coupled template and aggregate grow exponentially with the same exponent. Simulation of equations (2.6) and (3.4) verifies these findings (figure 2).

REFERENCES

Burdett, I. D. J. & Kirkwood, T. B. L. 1983 How does a bacterium grow during its cell cycle? *J. Theor. Biol.* **103**, 11–20. (doi:10.1016/0022-5193(83)90195-9)

Gánti, T. 2004 *Chemoton theory*. Dordrecht, The Netherlands: Kluwer Academic.

Lee, D. H., Granja, J. R., Martinez, J. A., Severin, K. & Reza Ghadri, M. 1996 A self-replicating peptide. *Nature* **382**, 525–528. (doi:10.1038/382525a0)

Luisi, P. L., Ferri, F. & Stano, P. 2006 Approaches to a semi-synthetic minimal cell: a review. *Naturwissenschaften* **93**, 1–13. (doi:10.1007/s00114-005-0056-z)

Mattes, A. & Seitz, O. 2001 Sequence fidelity of a template-directed PNA-ligation reaction. *Chem. Commun.* 2050–2051. (doi:10.1039/b106109g)

Munteanu, A., Attolini, C. S.-O., Rasmussen, S., Ziock, H. & Solé, R. V. 2007 Generic Darwinian selection in catalytic protocell assemblies. *Phil. Trans. R. Soc. B* **362**, 1847–1855. (doi:10.1098/rstb.2007.2077)

Nielsen, P. E. 1999 Peptide nucleic acid, a molecule with two identities. *Acc. Chem. Res.* **32**, 624–630. (doi:10.1021/ar980010t)

Novak, B., Csikasz-Nagy, A., Gyorffy, B., Chen, K. & Tyson, J. J. 1998 Mathematical model of the fission yeast cell cycle with checkpoint controls at the G1/S, G2/M and metha-phase/anaphase transitions. *Biophys. Chem.* **72**, 185–200. (doi:10.1016/S0301-4622(98)00133-1)

Paul, N. & Joyce, G. F. 2004 Minimal self-replicating systems. *Curr. Opin. Chem. Biol.* **8**, 634–639. (doi:10.1016/j.cbpa.2004.09.005)

Pohorille, A. & Deamer, D. W. 2002 Artificial cells: prospects for biotechnology. *Trends Biotechnol.* **20**, 123–128.

Pschl, A., Sforza, S., Haaima, G., Dahl, O. & Nielsen, P. E. 1998 Peptide nucleic acids (PNAs) with a functional backbone. *Tetrahedron Lett.* **39**, 4707–4710.

Rasmussen, S., Chen, L., Nilsson, M. & Abe, S. 2003 Bridging nonliving and living matter. *Artif. Life* **9**, 269–316. (doi:10.1162/106454603322392479)

Rasmussen, S., Chen, L., Deamer, D., Krakauer, D. C., Packard, N. H., Stadler, P. E. & Bedau, M. A. 2004 Transitions from nonliving to living matter. *Science* **303**, 963–965. (doi:10.1126/science.1093669)

Rosen, M. H. J. 1988 *Surfactants and interfacial phenomena*. New York, NY: Wiley Interscience.

Sievers, D. & Von Kiedrowski, G. 1994 Self-replication of complementary nucleotide-based oligomers. *Nature* **369**, 221–224. (doi:10.1038/369221a0)

Stadler, B. M. R. & Stadler, P. F. 2003 Molecular replicator dynamics. *Adv. Comp. Syst.* **6**, 47–77. (doi:10.1142/S0219525903000724)

Stadler, B. M. R., Stadler, P. F. & Schuster, P. 2000 Dynamics of autocatalytic replicator networks based on higher-order ligation reactions. *Bull. Math. Biol.* **62**, 1061–1086. (doi:10.1006/bulm.2000.0194)

Szathmáry, E. 2005 Life: in search of the simplest cell. *Nature* **433**, 469–470. (doi:10.1038/433469a)

Szostak, J. W., Bartel, D. P. & Luisi, P. L. 2001 Synthesizing life. *Nature* **409**, 387–390. (doi:10.1038/35053176)

Tyson, J. J., Chen, K. C. & Novak, B. 2003 Sniffers, buzzers, togglers, and blinkers: dynamics of regulatory and signaling pathways in the cell. *Curr. Opin. Cell Biol.* **15**, 221–231. (doi:10.1016/S0955-0674(03)00017-6)

Varga, Z. & Szathmáry, E. 1997 An extremum principle for parabolic competition. *Bull. Math. Biol.* **59**, 1145–1154. (doi:10.1007/BF02460105)

Molecular Dynamics Study of Small PNA Molecules in Lipid-Water System

Paweł Weroński,*† Yi Jiang,* and Steen Rasmussen‡§
*Theoretical Division and Center for Nonlinear Studies, and ‡Earth and Environmental Sciences Division, Los Alamos National Laboratory, Los Alamos, New Mexico; †Institute of Catalysis and Surface Chemistry, Polish Academy of Sciences, Kraków, Poland; and §Santa Fe Institute, Santa Fe, New Mexico

ABSTRACT We present the results of molecular dynamics simulations of small peptide nucleic acid (PNA) molecules, synthetic analogs of DNA, at a lipid bilayer in water. At neutral pH, without any salt, and in the $NP_n\gamma T$ ensemble, two similar PNA molecules (6-mers) with the same nucleic base sequence and different terminal groups are investigated at the interface between water and a 1-palmitoyl-2-oleoylphosphatidylcholine lipid bilayer. The results of our simulations suggest that at low ionic strength of the solution, both PNA molecules adsorb at the lipid-water interface. In the case where the PNA molecule has charged terminal groups, the main driving force of adsorption is the electrostatic attraction between the charged groups of PNA and the lipid heads. The main driving force of adsorption of the PNA molecule with neutral terminal groups is the hydrophobic interaction of the nonpolar groups. Our simulations suggest that the system free energy change associated with PNA adsorption at the lipid-water interface is on the order of several tens of kT per PNA molecule in both cases.

INTRODUCTION

Peptide nucleic acids (PNA) are lab-created analogs of DNA, in which the nucleic bases (adenine, guanine, thymine, and cytosine) are attached to a pseudopeptide backbone. PNA can hybridize to its complementary DNA target in a sequence-dependent manner (1). Unlike most oligonucleotide analogs, PNA binds very tightly to double-stranded DNA as well. Because of this property, PNA molecules are intensively exploited in gene chemistry. They provide a novel, straightforward, and versatile approach for permanently attaching proteins, peptides, fluorophores, and other molecules to plasmid DNA without interfering with transcriptional activity (2).

Recently, due to PNA's neutral backbone, a new potential application emerged. PNA or backbone modified lipophilic PNA can act as the genetic material in a minimal self-replicating nanomachine or protocell design proposed by Rasmussen and Chen (3,4). According to this design, a minimal protocell able to utilize resources, grow, self-replicate, and evolve could be as simple as a small lipid aggregate (micelle) acting as a container by anchoring a PNA molecule to its exterior and a photosensitizer to its interior. In such a protocell, light-driven metabolic processes from the lipid aggregate interior could synthesize lipids and PNA, from appropriate precursor molecules with PNA acting as both an information carrier and as a catalyst, leading to a spontaneous growth of the protocell. Micellar lipid clusters in water are thermodynamically stable below some critical size only; therefore the growing containers would divide as soon as they become large enough, providing the next generation of the protocells.

To verify this scenario we need to answer a number of important questions. Unlike DNA or RNA, our knowledge of PNA is still relatively limited and a number of important issues have not been addressed so far. One of them is whether PNA molecules at high local concentrations are going to spontaneously gather at a lipid-water interface, and if not, how hydrophobic the PNA backbone has to be before such an attachment occurs. The PNA-lipid attachment is a sine qua non condition for the protocell replication process. Molecular dynamics (MD) computer simulation is a suitable method for addressing this problem. This technique has been exploited with a significant success in many areas of bioscience and bioengineering, including some recent PNA-related phenomena (5,6). MD simulation is also one of the main tools used in theoretical studies of membranes and lipid bilayers (7). In this article we present results of our MD simulations on the affinity of a small PNA molecule to a lipid-water interface.

METHODS

General

All the MD simulations were conducted with NAMD, version 2.6b1 (8) using CHARMM27 topology and parameter files (9). We used VMD, version 1.8.3 (10), to prepare input files and to analyze the MD trajectories, as well as for visualization. All the calculations were performed on Mauve, the SGI Altix supercomputer of the Los Alamos National Laboratory (LANL) and on the Linux cluster of the Center for Nonlinear Studies at LANL.

Topology and parameterization

Unfortunately, the topology and parameters for PNA residues are not directly available in the standard molecular dynamics packages such as CHARMM (11) because PNA molecules are not natural or common. To the best of our knowledge, only two sets of parameters for PNA have been published in the literature (12–14). Both parameter sets, however, seem to be a little inconsistent with each other. More importantly, some of the parameters in these sets are different than parameters found in CHARMM27 parameter file for similar chemical structures (compare, e.g., partial atomic charges for the

Submitted September 12, 2006, and accepted for publication December 27, 2006.

Address reprint requests to Paweł Weroński, Theoretical Division, Los Alamos National Laboratory, MS-B284, Los Alamos, NM 87545. Tel.: 505-667-9956; Fax: 505-665-2659; E-mail: pawel@lanl.gov.

0006-3495/07/05/3081/11 $2.00 doi: 10.1529/biophysj.106.097352

peptide bond). Therefore, to be consistent with the other parameters used in our simulations, we have obtained the parameters in PNA residues by comparison with similar chemical groups found in the CHARMM27 parameter file (15–21). We have also assigned the atom types of the atoms involved in PNA residues as reported in Sen and Nilsson (13), to follow CHARMM atom type definitions (see Fig. 1). This effort makes most of the bond, angle, and dihedral and all of the nonbonded parameters directly available from the combined CHARMM all-hydrogen parameter set for nucleic acids, lipids, and proteins. These, which are not available directly, are listed in the Supplementary Material. Corresponding topology and parameters files for NAMD calculations are also available upon request. Our force-field parameters have been tested in ongoing simulations of PNA partition in water-octanol system and the experimentally verified results will be published elsewhere.

Construction and equilibration of the starting structures

We extracted all the atoms' coordinates of one palmitoyl-oleoyl-phosphatidyl-choline (POPC) molecule from the file fluid-H.pdb.Z that we downloaded from Dr. H. Heller's web page (22,23). The molecule was aligned with z axis and replicated 6×6 times in the directions of the axes x and y every 0.75 nm. This procedure produced a monolayer of 36 lipid molecules in the xy plane. The second compartment of the bilayer was produced by reflection of the monolayer in this plane. The two monolayers were separated with a gap of 4.5 nm. We used the SOLVATE facility of VMD to fill the gap with 2895 TIPS3P water molecules. The resulting system of the size $4.5 \times 4.5 \times 10.6$ nm was then energy-minimized for 500 time steps to avoid overlapping of atoms. After that, a short (4500 time steps) simulation was run at constant temperature, pressure, and constant shape of the simulation box in directions x and y (x/y ratio) to allow shape relaxation of the unit cell. Next we conducted an ~4-ns-long simulation at constant temperature, constant pressure component normal to xy plane, and constant area of the unit cell in this plane. At this stage of the simulation, the dimensions of the unit cell in the directions x and y were fixed at 4.743 nm that allowed the bilayer to expand to this size, which corresponds to the molecular area 0.625 nm^2 per lipid found in experiments (24). At the same time the size of the unit cell in the direction z decreased to the value of 7.75 nm. Finally, we ran several 40-ns simulations of the system at constant temperature, normal pressure, shape of the simulation box in directions x and y, and a few values of the surface tension. We found that at the surface tension $\gamma = 0.025\,\mathrm{N/m}$ the molecular area was $A_0 = 0.623 \pm 0.011\,\mathrm{nm}^2$, which fitted best the experimental result. Therefore, all the further simulations involving the lipid bilayer were conducted at this specific value of the surface tension.

The coordinates of the atoms in the single-stranded 6-mer PNA with the base sequence $C_1G_1TAC_2G_2$ were extracted from the crystal structure of the corresponding PNA duplex (Protein Data Bank identifier 1PUP (25)). The coordinates of the missing H-atoms of the PNA molecule were generated by the GUESSCOORD facility of VMD. We created two PNA molecules with the same base sequence and different terminal groups. One of them, hereafter referred to as s-PNA, has standard C- and N-terminus, i.e., the charged groups $-COO^-$ and $-NH_3^+$, respectively. The second molecule that we will call a-PNA has two neutral termini: amidated C-terminus $-NH_2$ and acetylated N-terminus $-CO-CH_3$ (see Fig. 2). Each of the molecules was solvated in water and equilibrated for 40 ns at constant temperature, pressure, and cross-section area. The size of the simulation boxes in the directions x and y was fixed at the value corresponding to the size of the lipid bilayer, i.e., at 4.743×4.743 nm. The z dimension of the box fluctuated around 3.9 nm.

After the equilibration, each of the two PNA molecules was inserted into the equilibrated lipid bilayer in the following way. First, the equilibrated lipid bilayer system was moved along z axis in such a way that the layer of water was located in the center of the simulation box. Next, the PNA molecule was centered in its simulation box. Then, we removed a volume of water from the center of the lipid bilayer system; the volume was 0.2 nm thicker in z direction than the size of the corresponding PNA molecule. Finally, the empty space was replaced with the PNA molecule together with its surrounding water molecules. The resulting system was then energy-minimized for 400 time steps to avoid overlapping of atoms and then used for the subsequent MD simulations.

It should be noted that our systems are different than experimental systems that have been studied so far, e.g., in the PNA vesicle encapsulation experiments (26). First of all, there is no salt or electrolyte in our systems and the lipid-water interface in our simulations is electrostatically neutral. In contrary, the experimental systems have usually been studied at the buffer concentration on the order of at least several millimolar and the measured surface potential of the lipid-water interface is on the order of 10^2mV (26). This large value of the interface surface potential suggests that strong ion adsorption or desorption can take place at the interface, which in turn can produce a compact hydration layer and influence the PNA molecule behavior at the interface. Indeed, preliminary results of our simulations of the lipid bilayer at ionic strength on the order of fraction of mol/dm^3 suggest that this is the case. It should also be noted that because of the MD length-scale limitation the PNA and lipid concentrations in our systems are equal to $\sim 10^{-2}$ M and 0.7 M, respectively, and so they are very high compared to the experimental conditions.

Molecular dynamics protocols

We employed NAMD (version 2.6b1) with its standard empirical potential energy function (8). We used periodic boundary conditions and the particle mesh Ewald (PME) approach (27,28) for evaluating long-range electrostatic effects. The distance between subsequent grid points on the mesh was ~0.1 nm. A time step of 2 fs was used for the integration of Newton's equations, since all the bond lengths between each hydrogen atom and the atom to which it was covalently bonded were constrained to their equilibrium values

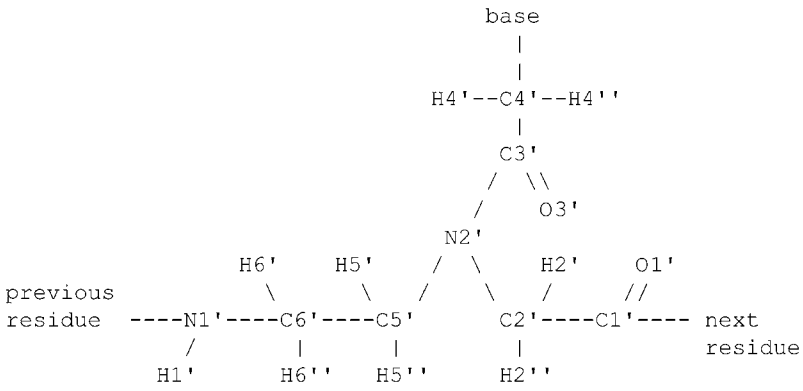


FIGURE 1 Schematic representation of the chemical structure of the monomeric unit of PNA.

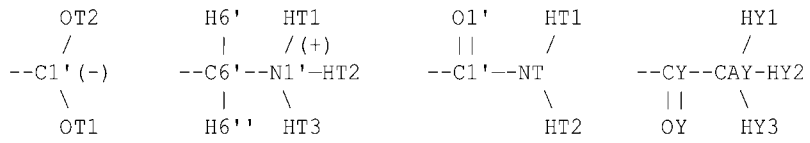


FIGURE 2 Schematic representation of the terminal groups used in our MD simulations (from *left to right*): standard C-terminus of the s-PNA molecule, standard N-terminus of the s-PNA molecule, amidated C-terminus of the a-PNA molecule, and acetylated N-terminus of the a-PNA molecule. Both PNA molecules contain the same base sequence ($C_1G_1TAC_2G_2$).

using the SHAKE algorithm (29). A cutoff at 1.2 nm was used for the Lennard-Jones interactions and for electrostatics with smoothing functions activated at a switching distance 1.0 nm. The nonbonded pair list was updated every 10 steps and the maximum distance between atoms for inclusion in the list was 1.35 nm. The long-range Coulombic forces were updated every two steps. For electrostatic calculations a relative dielectric constant of 1.0 was used; 1-2 and 1-3 nonbonded interactions were excluded; 1-4 electrostatic interactions were taken into account without any modification and 1-4 van der Waals parameters were modified according to the CHARMM27 parameter file.

Simulations were carried out in the ensemble $NP_n\gamma T$ at the constant pressure component normal to the lipid-water interface $P_n = 101.325\,\text{kPa}$, surface tension $\gamma = 0.025\,\text{N/m}$, and temperature $T = 298\,\text{K}$. Constant temperature was maintained by using the Langevin dynamics method (30). We applied the Langevin damping coefficient to nonhydrogen atoms only. Its value was setup to 5/ps for the first 4000 time steps, when our systems were far from equilibrium. Then the coefficient was fixed at 1/ps. A combination of the Nose-Hoover constant pressure method (31) with piston fluctuation control implemented using Langevin dynamics (32) was used to control the pressure along the z axis and the surface tension in the xy plane. The Langevin piston oscillation period and the oscillation decay time were equal to 100 fs and 50 fs, respectively. The constant ratio of the unit cell dimensions in the xy plane was kept during the simulations while allowing cell shape fluctuations along all axes. The pressure was calculated using the hydrogen-group-based pseudomolecular virial and kinetic energy (useGroupPressure option of NAMD) in conjunction with the SHAKE algorithm.

Calculation of the free energy

We calculated the free energy change in our systems using several different methods. The first method was based on the adaptive biasing force (ABF) approach (33). This method, implemented as a suite of Tcl routines directly available from the main configuration file used to run molecular dynamics simulations with NAMD, efficiently calculates the potential of mean force $\langle F_\xi\rangle_\xi$ acting between two groups of atoms along a reaction coordinate ξ, and applies ABFs that are needed to overcome free energy barriers and provide uniform sampling. This potential of the mean force is a measure of the system's free-energy change ΔG corresponding to the change in the reaction coordinate of a system from ξ_1 to ξ_2:

$$\Delta G = -\int_{\xi_1}^{\xi_2} \langle F_\xi\rangle_\xi \,\mathrm{d}\xi \cong -\sum_{\mathrm{i}} \langle F_\xi^{\mathrm{i}}\rangle_\xi \,\Delta\xi_{\mathrm{i}}, \tag{1}$$

where $\Delta\xi_{\mathrm{i}}$ is the width of the bin over which the force is averaged.

As discussed in the section ''Mean force and the free energy profile'', however, results of our simulations suggest that any biasing force acting on a flexible molecule like PNA or molecular structure such as the lipid bilayer can disturb its conformation and thus cause additional change of the free energy of the system. Therefore, to avoid the effect of the biasing force on the free energy calculation, we conducted our simulations monitoring the system's free energy without introducing any bias (applyBias option of NAMD).

The second method is strictly applicable to a system at equilibrium. Using the simulation data we can calculate the time evolution of the reaction coordinate and find the density of the distribution of this coordinate:

$$\rho(\xi) = \frac{1}{n_0}\frac{\mathrm{d}n(\xi)}{\mathrm{d}\xi} \cong \frac{1}{n_0}\frac{\Delta n_{\mathrm{i}}}{\Delta\xi_{\mathrm{i}}}, \tag{2}$$

where n_0 is the total number of the saved microstates of the system, $n(\xi)$ is the number of the microstates of the system in which the value of the reaction coordinate is smaller than ξ, and Δn_{i} is the number of the force measurements in the bin $\Delta\xi_{\mathrm{i}}$. The last equality is correct for large numbers Δn_{i} and small width of the bins $\Delta\xi_{\mathrm{i}}$. If the system is near equilibrium we can use the distribution density of the reaction coordinate ξ to calculate the change of the system free energy associated with the change of this coordinate. According to the Boltzmann relationship

$$\rho(\xi_2) = \rho(\xi_1)\exp[-\Delta G(\Delta\xi)], \tag{3}$$

where $\Delta\xi = \xi_2 - \xi_1$ is the change of the reaction coordinate and the free energy is expressed in kT units. The change of the system free energy ΔG corresponding to the change in the reaction coordinate of the system from ξ_1 to ξ_2 can be calculated from Eq. 3 as

$$\Delta G(\Delta\xi) = -\ln\frac{\rho(\xi_2)}{\rho(\xi_1)}. \tag{4}$$

If the width of the bin $\Delta\xi_{\mathrm{i}}$ is constant, the change of the system free energy can be calculated from the formula

$$\Delta G(\Delta\xi) \cong -\ln\frac{\Delta n_{\mathrm{i}2}}{\Delta n_{\mathrm{i}1}}, \tag{5}$$

where $\Delta n_{\mathrm{i}1}$ and $\Delta n_{\mathrm{i}2}$ are the numbers of force measurements collected in the bins corresponding to the reaction coordinates ξ_1 and ξ_2, respectively. Equation 5 is convenient because the numbers Δn_{i} are directly available in ABF calculations with NAMD. Note that if the free energy changes by more then few kT units in the whole reaction-coordinate interval of interest then the ABF simulation with no biasing force applied will take a very long time. To avoid this inconvenience the whole interval of reaction coordinate should be divided into smaller pieces, in which the free energy varies little, with a harmonic bias enforced at their borders (forceConst parameter of NAMD).

As the reaction coordinate we chose the distance along the axis z between a PNA molecule and one of the lipid-water interfaces. Specifically, we calculated the z coordinate of the PNA molecule by averaging the z coordinates of its selected 14 atoms. These were six atoms of the nucleic bases (O4, N4, N6, and O6 in the residues thymine, cytosine, adenine, and guanine, respectively), six atoms N2′ of the backbone, and two atoms of the terminal groups: N1′ and C1′ of the s-PNA molecule, and NT and CAY of the a-PNA molecule. The z coordinate of the lipid-water interface was calculated by averaging the z coordinates of the 36 N atoms and the 36 P1 atoms in this interface. The mean force was calculated between the 14 atoms of the PNA molecule and the 72 atoms of the lipid bilayer. The width of the bins in which the forces are accumulated has to be small enough to ascertain that the free energy profile varies regularly in the ξ-interval. On the other hand, it should be large enough to ensure sufficient sampling and avoid large fluctuations in the average force. In our system we used the bin width equal to 10 pm, which allowed us to collect the number of force samples in most of the bins on the order of 10^5–10^6. To decrease the force fluctuation we averaged the mean force over 20 pm (dSmooth parameter of NAMD).

As discussed in the next section, the molecule s-PNA adsorbs rapidly and forms hydrogen bonds with the charged groups of the lipid molecules that

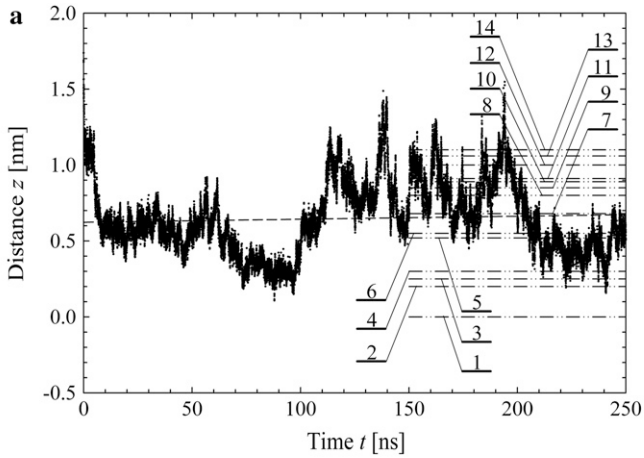
can be easily identified. In this system we can use yet another approach to estimate the PNA adsorption free energy. We can estimate the change of the system free energy by summing up the free energy of these bonds. When the s-PNA molecule adsorbs at the interface during the first nanosecond of the simulation and its distance from the interface does not evolve in a systematic way, we can assume that the system is at equilibrium after the first several nanoseconds and use Eq. 4 to calculate the free energy of each of the bonds. In these calculations we can assume that the lengths of the formed hydrogen bonds are uncorrelated and use the distance between the atoms forming a specific hydrogen bond as the reaction coordinate.

SIMULATION RESULTS AND DISCUSSION

In the following sections we present and discuss the results of the MD simulations of the s-PNA and a-PNA molecules in the lipid-water system: i), the variation of the distance between the molecules and the lipid-water interface; ii), the variation of the lipid molecular area; and iii), the molecule-interface free energy profile.

PNA distance from the lipid interface

The time evolution of the mean distance of the s-PNA molecule from the interface is presented in Fig. 3 *a* and in Fig. 3 *b* the final configuration is shown at $t = 250$ ns. The distance fluctuates around the mean value $\langle z \rangle = 0.68 \pm 0.24$ nm. No systematic drift of the distance can be observed in the timescale of the conducted simulation. From the simple linear regression analysis conducted over the entire time of the simulation we find that the slope of the dashed regression line is very small and equal to 0.20 ± 0.02 mm/s. The molecule approaches the interface quickly (in the first nanosecond of



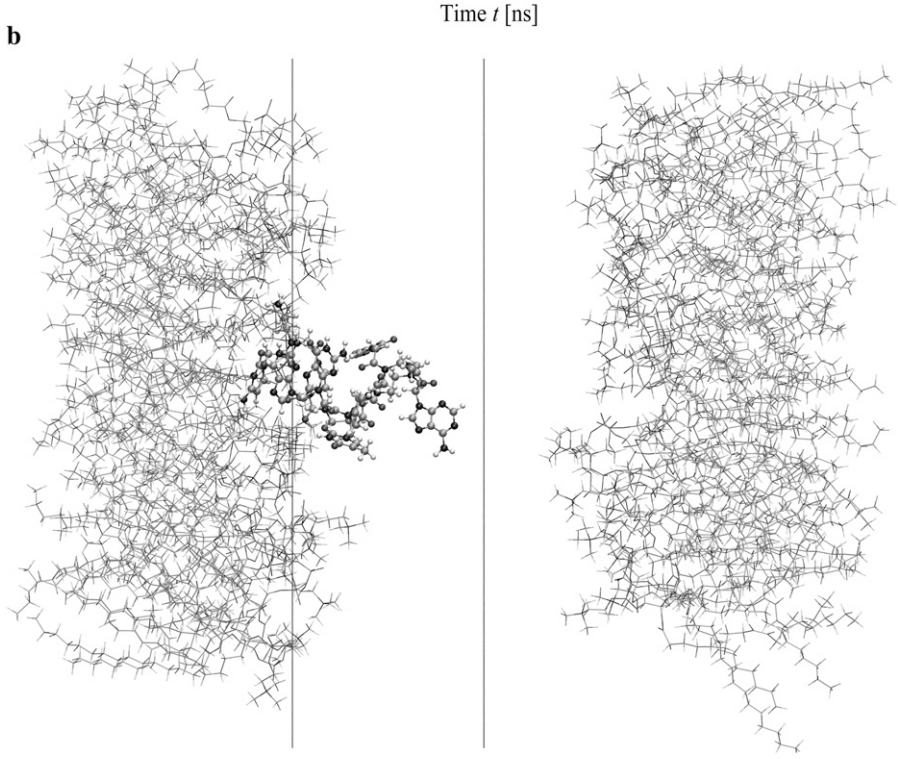


FIGURE 3 (*a*) Variation of the distance z between the lipid-water interface and the molecule s-PNA. Dots represent the distance between the molecule and the interface. The dashed lines represent the least square fitting of a straight line to the simulation results over the total time of the simulation. Dash-dot-dot lines represent the mean values of the distance z between the interface and the PNA molecule as well as between the interface and the 14 individual atoms of the molecule calculated during the last 100 ns of the simulations. Numbers correspond to the individual atoms of the molecule listed in the order of growing distance from the interface: 1 – N1′ (C_1), 2 – N4 (C_1), 3 – N2′ (C_1), 4 – O6 (G_1), 5 – N2′ (G_1), 6 – N4 (C_2), 7 – center of the molecule, 8 – N2′ (T), 9 – O6 (G_2), 10 – N2′ (C_2), 11 – N2′ (A), 12 – O4 (T), 13 – N6 (A), 14 – C1′ (G_2) and N2′ (G_2). (*b*) The final configuration of the system. The left and right vertical lines depict the coordinate $z = 0$ and $z = 1.8$ nm, respectively.

the simulation) and does not leave its position at the interface. Dash-dot-dot lines denote the mean distances of the 14 individual atoms of the molecule from the interface. The values of the mean distance and its standard deviation are listed in Table 1. Based on these data we can conclude that the molecule s-PNA is attached to the interface with its positively charged N-terminus, whereas the negatively charged C-terminus is pushed away from the interface.

Such an orientation of the molecule suggests that the attraction between the phosphate group of the POPC molecule and the N-terminus is stronger than that between the saturated amine group of the lipid molecule and the C-terminus of the s-PNA molecule. This difference in the attractive interaction results from the difference in the partial atomic charges of the attracting atoms, which are: $0.25e$ for each of the hydrogen atoms H11 through H43 of the saturated amine group, $-0.67e$ for each of the oxygen atoms OT1 and OT2 of the PNA C-terminus, $0.33e$ for each of the hydrogen atoms HT1 through HT3 of the PNA N-terminus, and $-0.78e$ for each of the oxygen atoms O3 and O4 of the lipid phosphate group, where e is the elementary proton charge. Note that the product of the oxygen-hydrogen partial atomic charges for the atoms of the N-terminus and the phosphate group is equal to $-0.2574e^2$, whereas for the atoms of the amine group and the C-terminus this product is equal to $-0.1675e^2$, which is 65% of the former value. This large difference between the products of the partial atomic charges suggests that the dipole-dipole interaction between the amine group and the C-terminus is weaker than that between the phosphate group and the N-terminus. Therefore, the hydrogen bonds between the lipid phosphate group and the N-terminus of the PNA molecule are energetically more favorable than those between the lipid amine group and the C-terminus of the PNA.

TABLE 1 Mean distance from the interface and its standard deviation calculated for individual atoms and the geometrical center of the molecule s-PNA

Atom name and residue	Mean distance from interface	Standard deviation	Position from interface
N1′ (C_1)	0.0	0.4	1
N2′ (C_1)	0.25	0.36	3
N4 (C_1)	0.2	0.4	2
N2′ (G_1)	0.52	0.33	5
O6 (G_1)	0.3	0.6	4
N2′ (T)	0.80	0.27	8
O4 (T)	1.0	0.4	12
N2′ (A)	0.91	0.26	11
N6 (A)	1.06	0.43	13
N2′ (C_2)	0.89	0.31	10
N4 (C_2)	0.55	0.45	6
N2′ (G_2)	1.1	0.4	14
O6 (G_2)	0.85	0.73	9
C1′ (G_2)	1.1	0.4	14
Center of molecule	0.68	0.24	7

The distance was averaged over the final stage of the simulation (100 ns). The numbers in the last column represent the position of the atoms in the order of growing distance from the interface.

The time evolution of the mean distance of the a-PNA molecule from the interface is presented in Fig. 4 *a* and the final configuration is shown in Fig. 4 *b* at $t = 350$ ns. Unlike the s-PNA molecule, the a-PNA molecule approaches the interface relatively slowly. Between 100 and 120 ns the molecule is even detached from the interface. We note a systematic drift of the mean distance of the a-PNA molecule from the interface toward smaller values. Slope of the dashed regression line calculated from the linear regression analysis of the simulation results is equal to -2.433 ± 0.012 mm/s. This large and negative value of the slope suggests that the molecule slowly sinks into the interface. Therefore, to achieve the equilibrium conformation of the system, we continued this simulation for a longer time until we saw that the free energy profile did not change any more, as discussed below. The mean distance of the a-PNA molecule from the interface, calculated in the time interval between 250 and 350 ns, fluctuates around the value $\langle z \rangle = 0.13 \pm 0.12$ nm. Thus, the distance is smaller by ~0.5 nm than that of the s-PNA molecule. Dash-dot-dot lines denote the mean values of the distance between the 14 individual atoms of the a-PNA molecule and the interface. The distances and their standard deviations are listed in Table 2. Based on these data we can conclude that the a-PNA molecule is attached to the interface with its hydrophobic groups, namely the acetylated N-terminus and amidated C-terminus. The smallest value of the z coordinate is that of the atom O4 of thymine. The collected frames of the simulation suggest that this is because there is a hydrogen bond between this atom and the N-terminus of the a-PNA molecule. Thus, the terminal groups of the molecule as well as the thymine base in the central part of the molecule sink into the interface. Therefore we can conclude that the main driving force of a-PNA adsorption is the hydrophobic interaction. Note that all values of the standard deviation listed in Table 2 are smaller than those in Table 1, which suggests that although the a-PNA adsorption process is slower than that of s-PNA, it eventually results in a stronger attachment of the molecule to the interface.

Molecular area per lipid

The time evolutions of the molecular area per lipid molecule, calculated for the systems with s-PNA and a-PNA, are presented in Fig. 5, *a* and *b*, respectively. The mean values of the molecular area are equal to $0.610 \pm 0.025\text{nm}^2$ and $0.631 \pm 0.023\text{nm}^2$, respectively. The calculated mean values of the molecular area deviate only slightly from the experimentally found value 0.625 nm^2 as well as from the value 0.623 found in our simulations without the PNA molecules. This means that introducing the PNA molecules does not change the surface tension of the lipid bilayer. Note that the mean value of the molecular area per lipid molecule is larger in the system with the a-PNA molecule by ~3% than that for the system with s-PNA. This may result from sinking of the hydrophobic parts of the a-PNA molecule into the internal

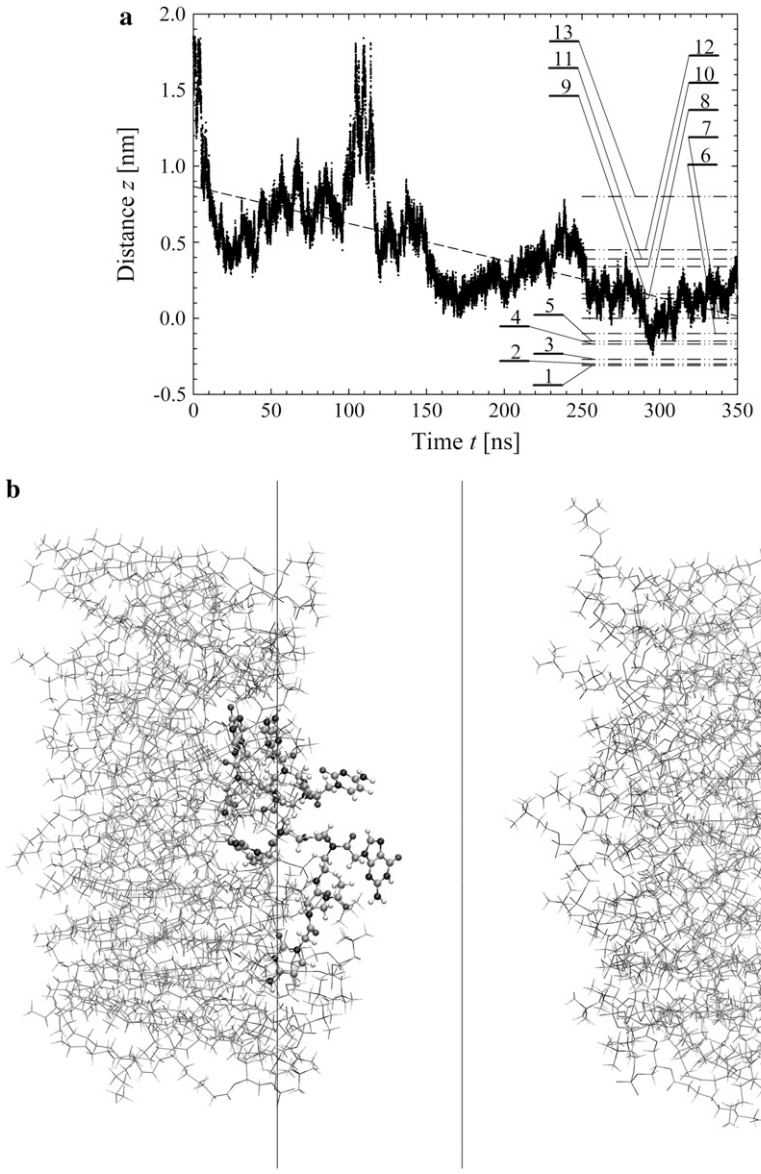


FIGURE 4 Same as Fig. 3 but for the molecule a-PNA. 1 – O4 (T), 2 – NT (G_2), 3 – N2′ (G_2), 4 – N4 (C_1), 5 – O6 (G_2), 6 – CAY (C_1), 7 – N2′ (C_1), 8 – center of the molecule, 9 – N2′ (T) and N6 (A), 10 – N2′ (C_2), 11 – N2′ (A), 12 – N2′ (G_1), 13 – O6 (G_1) and N4 (C_2). Note the systematic drift of the distance z toward the smaller values indicating that the hydrophobic parts of the a-PNA molecule slowly sink into the interface.

part of the bilayer, as discussed in the previous section. The difference between both molecular areas, however, is below the calculated standard deviation. Therefore longer simulations would be necessary to verify this hypothesis.

Mean force and free energy profile

The mean force and free energy profiles of our systems at the end of the simulations for the s-PNA and a-PNA molecules are presented in Fig. 6, *a* and *b*, respectively. Stars depict the calculated values of the mean force, whereas open and solid circles denote the values of the free energy calculated using Eqs. 1 and 4, respectively. The reference value of the free energy in all the profiles was set at the value corresponding to the distance 1.8 nm. Results obtained with both methods are very similar, especially in the case of a-PNA. Because of the rapid adsorption of the s-PNA molecule the sampling at the distance >1.5 nm is not sufficient. That is why at this range of the distance the fluctuations of the mean force are large and the free energy profile is inaccurate. Consequently, the free energy profiles in Fig. 6 *a* calculated with Eqs. 1 and 4 are shifted away from each other by $\sim 2\ kT$ units. Therefore, it is difficult to predict an exact depth of the free energy minimum based on the results obtained in our simulations.

The profile calculated for a-PNA is smoother. However, the configurations of our systems saved during the simulations suggest that once the PNA molecule adsorbs to the interface, the force measured between the molecule and the interface corresponds rather to a stretching of the molecule and the lipid bilayer than to detachment of the molecule from the interface, as the latter process is very rare. Therefore, we cannot estimate the value of the free energy change associated

TABLE 2 Mean distance from the interface and its standard deviation calculated for individual atoms and the geometrical center of the molecule a-PNA

Atom name and residue	Mean distance from interface	Standard deviation	Position from interface
CAY (C_1)	−0.1	0.3	6
N2′ (C_1)	0.0	0.2	7
N4 (C_1)	−0.17	0.32	4
N2′ (G_1)	0.45	0.14	12
O6 (G_1)	0.8	0.3	13
N2′ (T)	0.16	0.16	9
O4 (T)	−0.31	0.24	1
N2′ (A)	0.39	0.15	11
N6 (A)	0.16	0.16	9
N2′ (C_2)	0.34	0.17	10
N4 (C_2)	0.8	0.3	13
N2′ (G_2)	−0.27	0.15	3
O6 (G_2)	−0.15	0.21	5
NT (G_2)	−0.3	0.2	2
Center of molecule	0.13	0.12	8

The distance was averaged over the final stage of the simulation (100 ns). The numbers in the last column represent the position of the atoms in the order of growing distance from the interface.

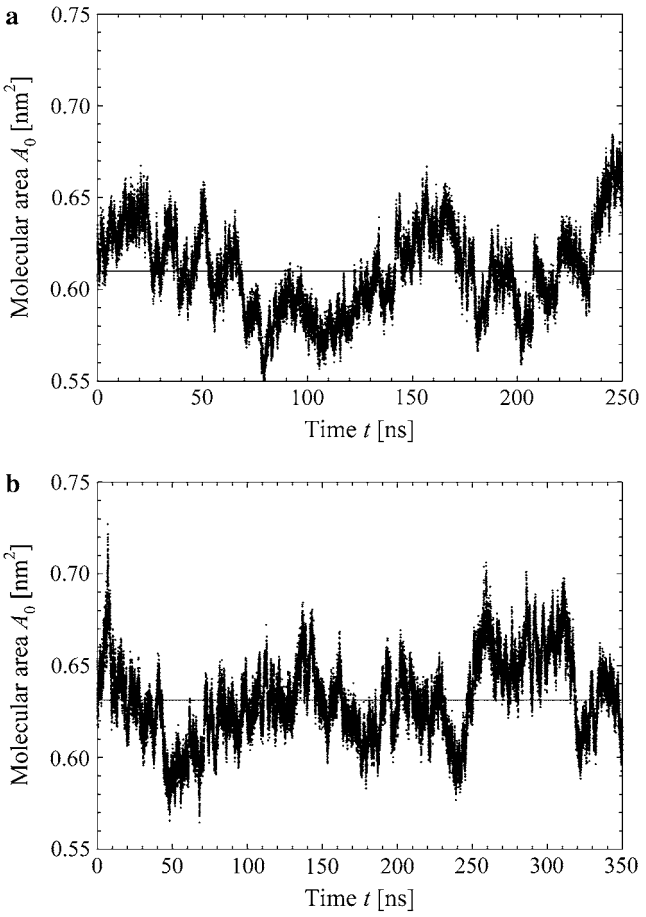


FIGURE 5 Time evolution of the molecular area per lipid calculated for the system containing the molecule s-PNA (*a*) and a-PNA (*b*). Solid lines denote the average values of the molecular area.

with adsorption of the a-PNA molecule from the data presented in Fig. 6 *b*.

For a comparison, two free energy profiles calculated with Eq. 1 and nonzero biasing force are presented in Fig. 6 *b* as well. The profiles correspond to adsorption and desorption of a-PNA molecule. Both profiles are quite different from the profiles obtained with no biasing force applied. The adsorption free energy profile calculated with the biasing force is shifted toward the larger distance z by ~0.5 nm. This shift is a direct consequence of pushing a-PNA molecule toward the interface, which does not allow a sufficient equilibration of the lipid bilayer. Therefore, the rapid increase of the adsorption free energy below $z < 0.5$ represents mainly the change of the system free energy corresponding to the deformation of the lipid bilayer under nonequilibrium conditions. The desorption free energy, on the other hand, increases monotonically with the distance z. This change of the system free energy results from a stretching of a-PNA molecule mainly and is not a good estimate of the PNA adsorption free energy as well.

The results obtained with the applied biasing force suggest therefore that the ABF method is not well suitable for calculating the system free energy change in the case of flexible molecules or molecular structures in general. In its classical formulation, the method can offer an accurate estimation of desorption energy only if there is a single, well-defined bond between the molecule and the interface, which can be chosen as the reaction coordinate. If there are a number of bonds between the molecule and the interface, however, calculating the free energy of a single bond will inevitably result in a deformation of the molecule and in an additional change of the system free energy. The applicability of the method can be improved through additional harmonic constraints applied to the flexible parts of the simulated system (P. Weroński and Y. Jiang, unpublished data).

As discussed above, the results presented in Fig. 3 *a* suggest that the main driving force of s-PNA adsorption at the lipid-water interface is the electrostatic attraction between these atoms of the POPC molecules and s-PNA molecule, which bear high partial atomic charges and form hydrogen bonds. Therefore, to estimate the s-PNA adsorption free energy more accurately, we used the approach based on the summation of the free energy of the hydrogen bonds formed between the PNA molecule and the interface. We used this method to calculate the depth of the free energy primary minima for selected 14 atoms of the first two residues, C_1 and G_1, which approach the lipid-water interface the most. We chose the atoms identified as hydrogen bond donors or acceptors: N1′ that is the N-terminus nitrogen atom, O1′, O2, N3, N4, and O3′ of the cytosine; and N2, O3′, N3, O6, N7, O1′, H1′, and H1 of the guanine.

Note that in case of the atoms N4 of the cytosine and N2 of the guanine, which are covalently bonded to more than one hydrogen atom, we calculated the distribution density of the

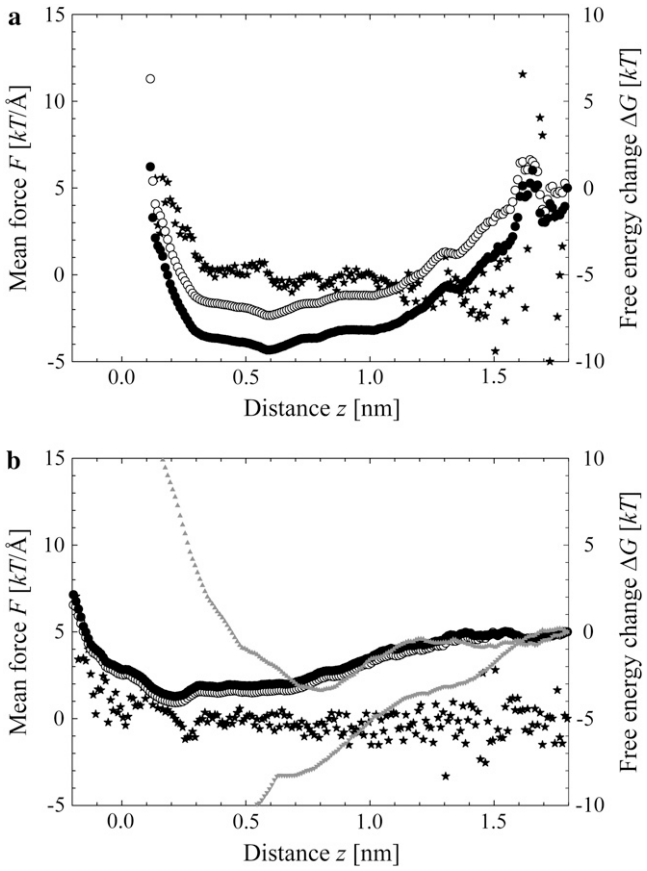


FIGURE 6 Free energy profile and mean force between the lipid-water interface and the molecule s-PNA (*a*) and a-PNA (*b*). Stars denote the mean force, open circles the free energy change calculated using Eq. 1, and solid circles depict the free energy change calculated using Eq. 5. Note large fluctuations of the force and free energy in the plot (*a*) above the distance $z = 1.6$ nm. Gray triangles in panel *b* denote the free energy profiles calculated with Eq. 1 when ABF is applied. These two profiles correspond to adsorption (▲) and desorption (▼) of a-PNA molecule.

distances between the nitrogen atoms and the nearest oxygen atom of the phosphate groups. In case of atoms N1 and N1′ of guanine, which are covalently bonded to single hydrogen atoms, we calculated the distribution density of the distances between these hydrogen atoms, H1 and H1′, and the nearest oxygen atom O3 or O4 instead. The other nine of the chosen 14 atoms, namely O2, N3, O1′, and O3′ of the cytosine; and N3, O6, N7, O1′, and O3′ of the guanine; formed hydrogen bonds with the hydrogen atoms of the lipid amine groups. For these atoms, the distribution density of the distance between them and the nearest hydrogen atom of the POPC amine groups was calculated.

The time evolution of the distance $r_{N1'C1}$ between the nitrogen atom N1′ of the cytosine C_1, to which the terminal hydrogen atoms HT1 through HT3 are bonded, and the nearest of the oxygen atoms O3 and O4 of the POPC molecules is presented in Fig. 7 *a*. The distribution density $\rho(r_{N1'C1})$ calculated with Eq. 2 for our system at $n_0 = 31250$ is presented in Fig. 7 *b*. We chose the intervals of the distance $r_{N1'C1}$ in such a way to keep the number of the microstates constant and equal to $\Delta n_i = 400$ except the last interval, where $\Delta n_{79} = 50$. The values of the distance $r_{N1'C1}$ reported in Fig. 7 *b* represent the average values calculated for each of the intervals.

The change of the system free energy calculated with Eq. 4 is presented in Fig. 7 *c*. We used the distribution density at the largest distance $r_{N1'C1} = 1.4$ nm as the reference value. Fig. 7 *c* shows that the depth of the primary minimum at the distance $r_{N1'C1} = 0.26$ nm, when the N-terminus nitrogen atom is separated from the nearest oxygen atom with a single hydrogen atom of the N-terminus, is equal to about -7 *kT*. Fig. 7 *c* also shows that, in addition to the primary well, the free energy profile exhibits three successive minima of decreasing values, located at the distance $r_{N1'C1}$ equal to 0.46 nm, 0.70 nm, and 0.91 nm. These minima correspond to the system configurations in which the N-terminus is separated from the lipid-water interface with one, two, and three water molecules.

The results of our calculations are listed in Table 3. The deepest energy minimum of the value -7 *kT* is related to the hydrogen bonds formed by the terminal hydrogen atoms HT1 through HT3, bonded to the atom N1′ of the residue C_1. The adsorption energy of the s-PNA molecule estimated as a sum of the contributions originating from the individual atoms is equal to about -64 *kT* units, which is one order of magnitude larger than the values calculated from the free energy profiles in Fig. 6 *a*. We believe that this value is a better approximation of the actual PNA adsorption energy.

The PNA-bilayer interaction is very short ranged. The mean radii of the PNA molecules, calculated in the systems without a bilayer during equilibration of the PNA molecules in water, were equal to 0.7 ± 0.3 nm and 0.78 ± 0.34 nm for the molecule with charged and neutral termini, respectively. The smaller mean radius of the s-PNA molecule could result from the attractive interaction between its charged termini, which can make the molecule more compact. The data presented in Fig. 6, *a* and *b*, suggest that the interaction range of the PNA molecules and the lipid-water interface is comparable to the mean size of the molecules. It seems to be a little larger in the case of the molecule with charged termini, which is consistent with the fact that the electrostatic or molecular dipole interaction is longer ranged than the hydrophobic interaction.

The mean force and energy profiles changed during the simulations. In the case of the a-PNA molecule the change was caused by slow sinking of the molecule into the bilayer, as discussed above. Therefore, one can see that the free energy profile calculated for the a-PNA molecule extends to smaller values of the molecule-interface distance.

CONCLUSIONS

We use all-atom MD simulations to investigate adsorption of small PNA molecules at a lipid bilayer. Our results suggest that in a lipid-water system at low ionic strength PNA

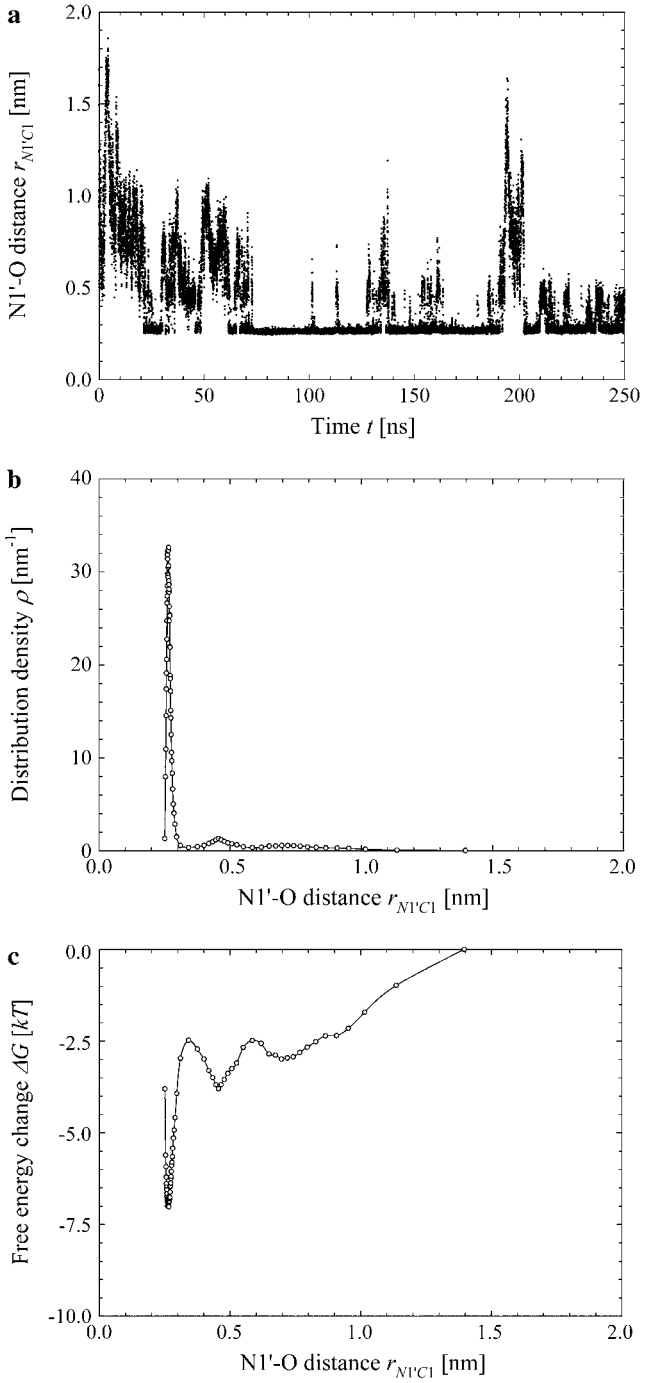


FIGURE 7 Simulation results for the hydrogen bond between the hydrogen atoms of the standard N-terminus and the oxygen atoms of the lipid phosphate groups. (*a*) Time evolution of the distance $r_{N1'C1}$ between the terminal nitrogen atom N1′ of the molecule s-PNA and the nearest oxygen atom O3 or O4 of the lipid phosphate groups; 250 ns of the time evolution is represented by 31,250 measurements of the distance $r_{N1'C1}$ at each 8 ps. The intervals $\Delta r_{N1'C1}$ are chosen in such a way to keep the number of the $r_{N1'C1}$ measurements per interval constant and equal to $\Delta n_i = 400$, except the last interval where $\Delta n_{79} = 50$. The reported values of the distance $r_{N1'C1}$ represent the average values calculated for each of the intervals. (*b*) Distribution density ρ of the distance $r_{N1'C1}$ between the terminal nitrogen atom N1′ of the molecule s-PNA and the nearest oxygen atom O3 or O4 of the lipid phosphate groups. The distribution density was calculated using Eq. 2. (*c*) Change of the system free energy ΔG as a function of the distance $r_{N1'C1}$ between the terminal nitrogen atom N1′ of the molecule s-PNA and the nearest oxygen atom O3 or O4 of the lipid phosphate groups. The free energy change was calculated with Eq. 4 and the distribution density at the largest distance $r_{N1'C1} = 1.4$ nm was used as the reference value. Note the deep primary minimum at the distance $r_{N1'C1} = 0.26$ nm, when the N-terminus nitrogen atom is separated from the nearest oxygen atom with a single hydrogen atom, as well as the three successive minima of decreasing values, located at the distance $r_{N1'C1}$ equal to 0.46, 0.70, and 0.91 nm, corresponding to the system configurations in which the N-terminus is separated from the lipid-water interface with one, two, and three water molecules.

TABLE 3 Partial atomic charge and hydrogen bond free energy calculated for six atoms of the residue C_1 and eight atoms of the residue G_1 of the molecule s-PNA

Mother atom name and residue	Hydrogen atoms names	Partial atomic charges (*e*)	Total bond free energy (*kT*)
N1′ (C_1)	HT1	0.33	−7.0
	HT2	0.33	
	HT3	0.33	
O1′ (C_1)	–	−0.51	−5.1
O2 (C_1)	–	−0.49	−4.8
N3 (C_1)	–	−0.66	−4.4
N4 (C_1)	H41	0.37	−3.5
	H42	0.33	
O3′ (C_1)	–	−0.51	−3.1
N2 (G_1)	H21	0.32	−5.1
	H22	0.35	
O3′ (G_1)	–	−0.51	−4.9
N3 (G_1)	–	−0.74	−4.8
O6 (G_1)	–	−0.51	−4.6
N7 (G_1)	–	−0.60	−4.5
O1′ (G_1)	–	−0.51	−4.1
N1′ (G_1)	H1′	0.31	−4.1
N1 (G_1)	H1	0.26	−3.8
Total ΔG			−63.8

The values of the free energy change were calculated using Eq. 4 and the interatomic distance distribution density found in the MD simulation that lasted 250 ns.

molecules spontaneously accumulate at the lipid-water interface. Considering that standard PNA molecules are typically soluble in water but not in organic solvents, it is interesting to find the PNA association with the lipid bilayer. We conclude that the change of the system free energy associated with adsorption of the small PNA molecules at the lipid-water interface is on the order of several tens of *kT* per PNA molecule consisting of six monomers. In the case of s-PNA molecules with charged terminal groups the main driving force of adsorption is the electrostatic attraction between the charged groups of the molecules and the charged heads of lipids. The absolute values of the partial atomic charges of the oxygen atoms of the POPC phosphate group, as well as the hydrogen atoms of the PNA N-terminal, are higher than the absolute values of the partial charges of the oxygen atoms of the PNA C-terminal and the hydrogen atoms of the lipid

amine group, respectively. Therefore, preferentially the positively charged terminus of the s-PNA molecule is bonded to the phosphate group of the POPC molecule, whereas the negatively charged C-terminus is pushed away from the interface. In the case of a-PNA molecules with neutral termini the main driving force of adsorption is the hydrophobic interaction of the nonpolar PNA groups. The hydrophobic parts of the a-PNA molecule sink into the organic interior of the bilayer, and therefore the mean distance of the molecule from the interface is smaller by a few angstroms compared to that of the s-PNA.

The PNA membrane association for low ionic strength and high concentration of the lipid and PNA is an interesting and perhaps even an unexpected prediction because PNA is known to solvate easily in water. Preliminary results of our simulations of the lipid bilayer at ionic strength and pH corresponding to the physiological conditions, however, suggest that strong adsorption of ions at the lipid-water interface takes place. Therefore, we conjecture that at higher ionic strength a compact hydration layer associated with the interface can effectively prohibit PNA adsorption, in agreement with the experimental results. At these conditions, due to the only weak adsorption of the standard PNA, we conjecture a backbone modified PNA could be more suited as a simple gene for the proposed protocell. Exactly how the backbone modifications need to be done has to be calculated and ultimately tested experimentally, where the presented results can be used as benchmarks. It should also be noted that the predicted PNA-lipid adsorption could play a role for the investigations based on PNA for gene therapy utilizing liposome delivery systems.

SUPPLEMENTARY MATERIAL

A supplement to this article can be found by visiting BJ Online at http://www.biophysj.org. The partial atomic charges for the atoms of the PNA backbone as well as the additional PNA parameters used in our simulations that are not provided in the standard CHARMM27 parameter files are available in the Supplementary Material.

We thank Prof. Lennart Nilsson for providing us with his topology and parameters files for PNA, Profs. Peter Nielsen and Ole Mouritsen for their comments and discussions of the simulation results, and all co-workers of the Protocell Assembly team at the Los Alamos National Laboratory for the many stimulating discussions and valuable suggestions.

This work was carried out under the auspices of the National Nuclear Security Administration of the U.S. Department of Energy at Los Alamos National Laboratory under contract No. DE-AC52-06NA25396.

REFERENCES

1. Nielsen, P. E., and M. Egholm. 1999. An introduction to peptide nucleic acid. *Curr. Issues Mol. Biol.* 1:89–104.
2. Zhilina, Z. V., A. J. Ziemba, and S. W. Ebbinghaus. 2005. Peptide nucleic acid conjugates: synthesis, properties and applications. *Curr. Top. Med. Chem.* 5:1119–1131.
3. Rasmussen, S., L. Chen, M. Nilsson, and S. Abe. 2003. Bridging nonliving and living matter. *Artif. Life*. 9:269–316.
4. Rasmussen, S., L. Chen, D. Deamer, D. C. Krakauer, N. H. Packard, P. F. Stadler, and M. A. Bedau. 2004. Transitions from nonliving to living matter. *Science*. 303:963–965.
5. Leach, A. R. 2001. Molecular Modelling: Principles and Applications, 2nd Ed. Prentice Hall, Harlow, UK.
6. Norberg, J., and L. Nilsson. 2002. Molecular dynamics applied to nucleic acids. *Acc. Chem. Res.* 35:465–472.
7. Gumbart, J., Y. Wang, A. Aksimentiev, E. Tajkhorshid, and K. Schulten. 2005. Molecular dynamics simulations of proteins in lipid bilayers. *Curr. Opin. Struct. Biol.* 15:423–431.
8. Kale, L., R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, K. Varadarajan, and K. Schulten. 1999. NAMD2: greater scalability for parallel molecular dynamics. *J. Comp. Phys.* 151:283–312.
9. CHARMM web site. http://www.charmm.org/document/Charmm/c31b1/index.html. Accessed February 10, 2005.
10. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: visual molecular dynamics. *J. Mol. Graph.* 14:33–38.
11. Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karpus. 1983. CHARMM: a program for macromolecular energy, minimization, and dynamics calculations. *J. Comput. Chem.* 4:187–217.
12. Shields, G. C., C. A. Laughton, and M. Orozco. 1998. Molecular dynamics simulation of a PNA·DNA·PNA triple helix in aqueous solution. *J. Am. Chem. Soc.* 120:5895–5904.
13. Sen, S., and L. Nilsson. 1998. Molecular dynamics of duplex systems involving PNA: structural and dynamical consequences of the nucleic acid backbone. *J. Am. Chem. Soc.* 120:619–631.
14. Sen, S., and L. Nilsson. 2001. MD Simulations of Homomorphous PNA, DNA, and RNA single strands: characterization and comparison of conformations and dynamics. *J. Am. Chem. Soc.* 123:7414–7422.
15. Schlenkrich, M., J. Brickmann, A. D. MacKerell, and M. Karplus. 1996. Empirical potential energy function for phospholipids: criteria for parameter optimization and applications. *In* Biological Membranes: A Molecular Perspective from Computation and Experiment. K. M. Merz and B. Roux, editors. Birkhauser, Boston, MA. 31–81.
16. MacKerell, A. D., D. Bashford, M. Bellott, R. L. Dunbrack, J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, et al. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B*. 102:3586–3616.
17. Feller, S. E., and A. D. MacKerell. 2000. Improved empirical potential energy function for molecular simulations of phospholipids. *J. Phys. Chem. B*. 104:7510–7515.
18. Foloppe, N., and A. D. MacKerell. 2000. All-atom empirical force field for nucleic acids. I. Parameter optimization based on small molecule and condensed phase macromolecular target data. *J. Comput. Chem.* 21:86–104.
19. MacKerell, A. D., and N. K. Banavali. 2000. All-atom empirical force field for nucleic acids. II. Application to molecular dynamics simulations of DNA and RNA in solution. *J. Comput. Chem.* 21:105–120.
20. Feller, S. E., K. Gawrisch, and A. D. MacKerell. 2002. Polyunsaturated fatty acids in lipid bilayers: intrinsic and environmental contributions to their unique physical properties. *J. Am. Chem. Soc.* 124:318–326.
21. Mackerell, A. D., M. Feig, and C. L. Brooks. 2004. Extending the treatment of backbone energetics in protein force fields: limitations of gas-phase quantum mechanics in reproducing protein conformational distributions in molecular dynamics simulations. *J. Comput. Chem.* 25:1400–1415.
22. Heller H. http://www.lrz-muenchen.de/~heller/membrane/fluid-H.pdb.Z. Accessed May 20, 2005.
23. Heller, H., M. Schaefer, and K. Schulten. 1993. Molecular dynamics simulation of a bilayer of 200 lipids in the gel and in the liquid-crystal phases. *J. Phys. Chem.* 97:8343–8360.
24. Vacklin, H. P., F. Tiberg, G. Fragneto, and R. K. Thomas. 2005. Phospholipase A_2 hydrolysis of supported phospholipid bilayers: a

neutron reflectivity and ellipsometry study. *Biochemistry.* 44:2811–2821.

25. Rasmussen, H., J. S. Kastrup, J. N. Nielsen, J. M. Nielsen, and P. E. Nielsen. 1997. Crystal structure of a peptide nucleic acid (PNA) duplex at 1.7 A resolution. *Nat. Struct. Biol.* 4:98–101.
26. Wittung, P., J. Kajanus, K. Edwards, G. Haaima, P. E. Nielsen, B. Norden, and B. G. Malmstrom. 1995. Phospholipid membrane permeability of peptide nucleic acid. *FEBS Lett.* 375:27–29.
27. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh Ewald: an N.log(N) method for Ewald sums in large systems. *J. Chem. Phys.* 98:10089–10092.
28. Essmann, U., L. Perera, M. L. Berkowitz, T. Darden, H. Lee, and L. G. Pedersen. 1995. A smooth particle mesh Ewald method. *J. Chem. Phys.* 103:8577–8593.
29. Ryckaert, J. P., G. Ciccote, and J. C. Berendsen. 1977. Numerical integration of the Cartesian equations of motion of a system with constraints: molecular dynamics of n-alkanes. *J. Comput. Phys.* 23: 327–341.
30. van Gunsteren, W. F., and H. J. C. Berendsen. 1982. Algorithms for Brownian dynamics. *Mol. Phys.* 45:637–647.
31. Martyna, G. J., D. J. Tobias, and M. L. Klein. 1994. Constant pressure molecular dynamics algorithms. *J. Chem. Phys.* 101:4177–4189.
32. Feller, S. E., Y. H. Zhang, R. W. Pastor, and B. R. Brooks. 1995. Constant pressure molecular dynamics simulation: the Langevin piston method. *J. Chem. Phys.* 103:4613–4621.
33. Henin, J., and C. Chipot. 2004. Overcoming free energy barriers using unconstrained molecular dynamics simulations. *J. Chem. Phys.* 121: 2904–2914.

ASTROBIOLOGY
Volume 6, Number 3, 2006

Hypothesis Paper

Experimentally Tracing the Key Steps in the Origin of Life: The Aromatic World

PASCALE EHRENFREUND,[1] STEEN RASMUSSEN,[2,3,4] JAMES CLEAVES,[5] and LIAOHAI CHEN[6]

ABSTRACT

Life is generally believed to emerge on Earth, to be at least functionally similar to life as we know it today, and to be much simpler than modern life. Although minimal life is notoriously difficult to define, a molecular system can be considered alive if it turns resources into building blocks, replicates, and evolves. Primitive life may have consisted of a compartmentalized genetic system coupled with an energy-harvesting mechanism. How prebiotic building blocks self-assemble and transform themselves into a minimal living system can be broken into two questions: (1) How can prebiotic building blocks form containers, metabolic networks, and informational polymers? (2) How can these three components cooperatively organize to form a protocell that satisfies the minimal requirements for a living system? The functional integration of these components is a difficult puzzle that requires cooperation among all the aspects of protocell assembly: starting material, reaction mechanisms, thermodynamics, and the integration of the inheritance, metabolism, and container functionalities. Protocells may have been self-assembled from components different from those used in modern biochemistry. We propose that assemblies based on aromatic hydrocarbons may have been the most abundant flexible and stable organic materials on the primitive Earth and discuss their possible integration into a minimal life form. In this paper we attempt to combine current knowledge of the composition of prebiotic organic material of extraterrestrial and terrestrial origin, and put these in the context of possible prebiotic scenarios. We also describe laboratory experiments that might help clarify the transition from nonliving to living matter using aromatic material. This paper presents an interdisciplinary approach to interface state of the art knowledge in astrochemistry, prebiotic chemistry, and artificial life research. Key Words: Origin of life—Aromatic hydrocarbons—Minimal life—Young Earth. Astrobiology 6, 490–520.

[1]Leiden Institute of Chemistry, Leiden, The Netherlands.
[2]Los Alamos National Laboratory, Los Alamos; and [3]Santa Fe Institute, Santa Fe, New Mexico.
[4]University of Copenhagen, IMBG Panum Institute, Kobenhavn, Denmark.
[5]Geosciences Research Division, Scripps Institution of Oceanography, La Jolla, California.
[6]Bioscience Division, Argonne National Laboratory, Argonne, Illinois.

INTRODUCTION

IT IS GENERALLY AGREED that life began relatively rapidly after conditions on the surface of the Earth allowed it, at least 3.85 billion years ago (Mojzsis *et al*., 1996), though the evidence from the geological record is still under debate (Nisbet and Sleep, 2001; van Zuilen *et al*., 2002). Because very few data are available regarding the atmospheric, oceanic, or geological conditions on the pre-biological Earth, it is difficult to determine conclusively whether terrestrially synthesized organic material or extraterrestrial molecules were more significant for life's origin. A fundamental question remains as to how simple molecules of terrestrial or extraterrestrial origin assembled into larger functional units under plausible geochemical conditions and then into replicating structures. Advances in directed molecular evolution, molecular self-assembly, and artificial redox and photochemical metabolic systems make the synthesis of protocells an imaginable goal (Szostak *et al*., 2001; Rasmussen *et al*., 2004).

There are two fundamental approaches to the study of the origin of life. One, the top-down approach, considers the origin of the common components of modern biochemistry and their organization. The other, the bottom-up approach, considers which compounds may have been plausibly produced under primitive planetary conditions and how they came to self-assemble. The top-down approach is biased by the uniformity of modern biochemistry across the three extant domains of life (Archaea, Bacteria, and Eukarya), which clearly originated from a common ancestor (Pace, 2001; Zhaxybayeva and Gogarten, 2004). Whatever the first living being was, the origin of life likely depended on the presence of organic compounds. Understanding the source and nature of this material is crucial for our understanding of subsequent molecular organization.

The endogenous synthesis of prebiotic organic compounds depends on the conditions on the young Earth (*e.g*., atmospheric composition, oxidation state of the early mantle, hydrothermal systems, glaciation, etc.), which are unfortunately poorly understood. Recent evidence from zircon crystals suggests that surface temperatures were relatively low and liquid oceans were present rather than a thick steam-rich atmosphere (Wilde *et al*., 2001). Such oceans could have facilitated the evolution of life. The amount of organic material produced by atmospheric synthesis or hydrothermal vent chemistry would have depended on the nature of the gases emitted by early volcanism. Extraterrestrial material is estimated to have delivered $\sim 10^9$ kg of carbon per year to the Earth during the heavy bombardment phase 4.5–3.8 billion years ago (Chyba and Sagan, 1992). For comparison, endogenous synthesis is estimated to have produced between $1/100^{\text{th}}$ this value for neutral atmospheres to 1,000 times this value for highly reducing atmospheres (Stribling and Miller, 1987). Ehrenfreund *et al*. (2002) provided a recent quantitative estimate of major sources of organic compounds on the early Earth. If the early atmosphere was not reducing, then infalling material from space may have been the predominant source. As our knowledge of the composition of interstellar dust and gas, comets, asteroids, and meteorites grows we can determine more accurately the molecular inventory of material that was transported to young planets by exogenous delivery. Recent astronomical observations and laboratory analysis of extraterrestrial material suggest that the majority of carbon in the universe is in the form of aromatic solid macromolecular material (Ehrenfreund *et al*., 2004). Interstellar carbon chemistry proceeds similarly throughout the universe as evidenced by astronomical observations of galactic and extragalactic regions (*e.g*., Spoon *et al*., 2003; Peeters *et al*., 2005; Sollerman *et al*., 2005). Therefore, extraplanetary input of organic matter may proceed in similar ways in other solar systems as well.

Life must have initially been fairly simple and may have used the most abundant and stable material present on the young Earth. Astronomical observations of comets and laboratory measurements of carbonaceous meteorites show that the most abundant carbon material they likely carried to Earth were aromatic molecules. Aromatic molecules compacted in macromolecular carbon or as free volatiles may have formed on the early Earth as well (Gold, 1999). Aromatic molecules are typically very stable. Highly aromatic macromolecular carbon (such as that found in meteorites) may fragment and release soluble organics upon chemical oxidation at slightly elevated temperatures. Components active in modern biochemistry are fragile with respect to heat, radiation, and oxidation, and therefore the self-assembly of life's precursor molecules from a pool of solid and volatile aromatic structures may be a more realistic scenario under hostile early Earth conditions.

In this paper we discuss the formation, inventory, and development of cosmic starting material for the origin of life in interstellar space (The Formation and Evolution of Organic Material in Space) and in the solar system (From the Interstellar Medium to the Terrestrial Planets). Terrestrial sources of organic matter and possible prebiotic scenarios on the young Earth are summarized in Terrestrial Sources of Organic Matter and Possible Prebiotic Scenarios. Despite many decades of laboratory efforts in prebiotic chemistry (using components present in contemporary life) no definitive origin of biochemistry scenario has been elaborated. In Functional Assemblies and Their Integration into a Minimal Life Form, we propose how assemblies of aromatic hydrocarbons might be integrated into a minimal life form. We discuss laboratory experiments that could provide insights into this hypothesis and suggest that the first terrestrial life may have been composed of compounds no longer used in modern biochemistry.

THE FORMATION AND EVOLUTION OF ORGANIC MATERIAL IN SPACE

To understand how life originated on Earth, it is useful to review several crucial steps in the evolution of the universe such as the formation of the chemical elements, molecules and cosmic dust, and the processes that led to the formation of planets. Observations of numerous protoplanetary disks and the detection of more than 170 extrasolar planets provide evidence that planets are common in the universe (see http://exoplanets.org/) (Marcy *et al*., 2005). The detection of small terrestrial planets will require further technological advances, though recently a frozen Earth-like planet, OGLE-2005-BLG-390Lb, has been discovered during a gravitational microlensing campaign (Beaulieu *et al*., 2006). Earth has provided unique conditions for life such as moderate surface temperatures, abundant water, a stable orbit and inclination, and a dynamic atmosphere and interior structure—properties that may be uncommon on other planets (Ehrenfreund and Martin, 2006). Organic chemistry in space, however, seems to follow common pathways throughout the universe. Carbonaceous molecules in the gas or solid state (carbonaceous dust or icy grains) are observed in similar abundances and composition in our and distant galaxies (Spoon *et al*., 2003; Peeters *et al*., 2005; Sollerman *et al*., 2005) (Fig. 1). In the last decade astronomical observations have allowed us to peer into distances of many billions of light years. Such observations of galaxies at high redshift probe the early universe and show that carbonaceous molecules were already present and dust formation already occurring (Bertoldi *et al*., 2003; Kaneda *et al*., 2005; Yan *et al*., 2005). Taking into account the dimensions of our universe, the ubiquitous observations of aliphatic and aromatic molecules, as well as C-bearing ices (such as CO), provide a clear indication that chemical processes have proceeded in similar ways. Such factors also compel us to consider Life in the larger context of astrophysics. The largest fraction of carbon in the universe is incorporated into solid aromatic macromolecular matter. Few studies have been performed with regard to how such material could have been involved in the origin of life. Insights into organic chemistry in space will assist our ability to predict whether life as we know it is likely to exist on other planets as well.

Formation of the elements

The universe is estimated to be approximately 13.7 billion years old. During Big Bang nucleosynthesis, only H and He and traces of a few other light nuclei such as D, T, Li, and Be were formed. Since the universe was expanding, cooling, and decreasing in density, heavier elements could not have formed during its early history. Carbon, the basis of organic chemistry, is produced by the so-called triple-α process ($3\ ^4\mathrm{He} \rightarrow\ ^{12}\mathrm{C}$) in the cores of stars. Today, stars more massive than half a solar mass form carbon after they have fused a significant fraction of their core hydrogen into helium. In the same cores, some carbon is further processed to the principal isotope of oxygen by $^{12}\mathrm{C} + {}^4\mathrm{He} \rightarrow\ ^{16}\mathrm{O}$. During the stellar burning cycle, contraction and heating of the stellar core alternate with a cycle of expansion and cooling. In this way, heavier elements (until ^{56}Fe) are produced in more massive stars. ^{56}Fe has the greatest nuclear binding energy and is unable to fuse with other nuclei to produce energy. The accumulation of ^{56}Fe in the stellar core destabilizes the star and triggers stellar collapse. All elements heavier than iron are formed during the final evolutionary period of stars dominated by mass loss, core degradation, and stellar explosions. The cosmic abundances of the chemical elements are summarized in Table 1. New perspectives on cosmic abun-

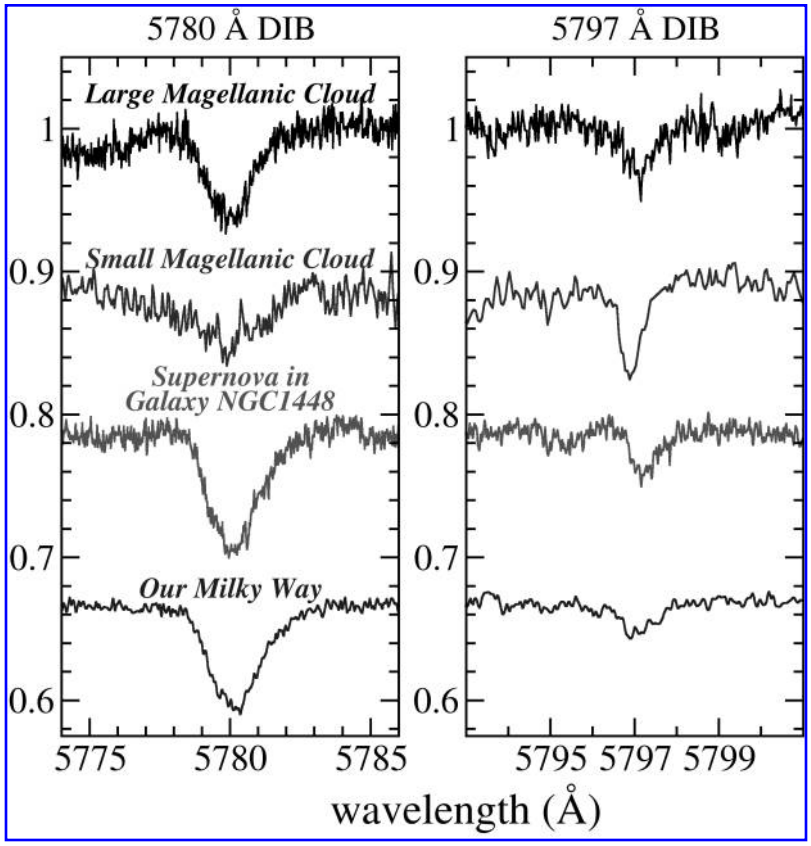


FIG. 1. More than 300 diffuse interstellar bands (DIBs) are observed in the absorption spectra of stars distributed throughout the universe. Those bands most probably originate from large carbonaceous molecules such as PAHs or fullerenes (Salama *et al.*, 1996). Two strong DIBs appear at 5780 and 5797 Å in the visible range. The signature of these bands can be observed in galactic and extragalactic regions. Here we display an example of those bands in our Milky Way, our neighbor galaxies the Large and Small Magellanic Clouds, and toward a supernova in a distant galaxy, NGC 1448.

dances have been recently summarized by Beers (2005).

The first stars

Recent cosmological data indicate that the formation of the first stars (population III non-metallic stars) may have occurred as early as 200 million years after the Big Bang (Daigne *et al.*, 2004). These very massive stars (probably more than 100 solar masses) were born from gravitationally contracting gas clouds dominated by hydrogen and helium in dark-matter halos (Scannapieco *et al.*, 2003). These stars can no longer be observed. The recent detection of hyper-metal-poor (HMPs) stars provides important constraints for early star formation processes (Iwamoto *et al.*, 2005). These HMP stars have a Fe/H ratio less than [1/100,000] of the solar ratio. They have an overabundance in C and O relative to Fe, and are slightly less massive than the Sun (Iwamoto *et al.*, 2005; Spaans and Silk, 2005). Investigations of those stars may provide us with clues with regard to the first generation of stars and supernovae that distributed the first heavy elements (Umeda and Nomoto, 2005) in the early universe.

The cycle of birth and death of stars constantly increases the abundance of heavy elements in the interstellar medium, a crucial prerequisite for terrestrial (rocky) planet formation and subsequently for the origin of life (for a review, see Spaans, 2004). In astronomical language, elements heavier than H and He are called "metals." Metals dispersed in the interstellar gas or incorporated into micron-sized dust particles have the ability to efficiently cool the interstellar gas.

TABLE 1. ELEMENTAL ABUNDANCES IN SPACE RELATIVE TO H

Element	*Abundance*
H	1
He	0.1
O	8.3×10^{-4}
C	4.0×10^{-4}
N	1.0×10^{-4}
S	1.7×10^{-5}
P	3.0×10^{-7}
Ne	0.8×10^{-4}
Si	4.3×10^{-5}
Mg	4.2×10^{-5}
Fe	4.3×10^{-5}
Na	2.1×10^{-6}

Those elements are excited by molecular collisions, and their return to stable energy levels releases energy *via* ultraviolet (UV) radiation in a much more efficient way than H collisions. In turn, cooling of the interstellar gas influences cloud fragmentation and triggers the formation of low-mass stars like our Sun (Scalo and Biswas, 2002). Low-mass stars have long life spans (*e.g.*, the Sun has a stable lifetime of ~10 billion years) that allow sufficient time to form terrestrial planets and life to evolve.

Understanding the early universe is extremely relevant to the question of terrestrial planet formation and consequently life. Observations and theory in the last decade indicate that star formation and heavy element enrichment occurred much earlier than previously thought (Daigne *et al.*, 2004).

From elements to molecules

Interstellar clouds and circumstellar envelopes act as nurseries for complex molecules and for siliceous and carbonaceous micron-sized dust particles. The interstellar medium constitutes ~10% of the mass of the galaxy. The interstellar medium can be subdivided into environments with very low-density hot gas, environments with warm intercloud gas, and regions with denser and colder material (Wooden *et al.*, 2004). H and He gas are the major components of interstellar clouds; molecules and submicron dust particles are only present in low concentration (Ehrenfreund and Charnley, 2000). Complex organic molecules can be formed through gas-phase reactions, solid-state chemistry, and gas–grain interactions. H_2 is by far the most abundant molecule in these clouds. CO is the most abundant carbon-containing species, with $CO/H_2 \sim 10^{-4}$. High excitation CO emissions from the most distant quasar currently known at redshift $z = 6.419$ have been detected, which indicates that CO has been present in the universe since approximately 800 millions years after the Big Bang (Bertoldi *et al.*, 2003). Silicate and carbon-based micron-sized dust particles, which are produced in the outflows of late-type stars, provide a catalytic surface for a variety of reactions when they are dispersed in molecular clouds (Ehrenfreund and Fraser, 2003). In cold clouds such dust particles adsorb ice mantles of water, CO_2, CO, and CH_3OH, with smaller admixtures of CH_4, NH_3, H_2CO, and HCOOH. Observations at infrared, radio, millimeter, and submillimeter frequencies show that a large variety of organic molecules are present in the dense interstellar gas (Charnley *et al.*, 2003). These include organics such as nitriles, aldehydes, alcohols, acids, ethers, ketones, amines, and amides, as well as long-chain hydrocarbons. The effective shielding of UV radiation by dust in such high-density regions enables complex gas–grain chemistry in the so-called "hot-core" regions close to the forming star, which allows the formation of many organic molecules, including precursor molecules for life as we know it (Kuan *et al.*, 2003) (see http://www.astrochemistry.net).

In addition to the rich inventory of prebiotic species observed in the gas phase in dense star-forming regions, laboratory experiments show the formation of amino acids upon UV irradiation of interstellar icy grain analogs (Bernstein *et al.*, 2002; Muñoz Caro *et al.*, 2002). Star-forming regions where interstellar icy grains evolve are strongly devoid of UV irradiation (Prasad and Tarafdar, 1983). Furthermore, it has been shown that prebiotic molecules such as amino acids and small heterocycles are rapidly degraded by UV radiation (Peeters *et al.*, 2003, 2005). The onset of the star-formation process is violent and destructive to any volatile or fragile (even when embedded in solid) material. There is, thus, no likely direct connection between simple interstellar prebiotic material (such as amino acids, sugars, etc.) in dense clouds and the origin of life on Earth. Observations of comets have shown that all of the original interstellar material has been strongly processed in the solar nebula (Ehrenfreund *et al.*, 2004). Only robust refractory material (such as silicates, metals, and solid carbonaceous matter) may retain significant interstellar heritage, as evidenced by isotopic enrichments (*e.g.*, Clemett *et al.*, 1993; Botta and Bada, 2002; Sephton *et al.*, 2004).

Carbon chemistry occurs most efficiently in circumstellar and diffuse interstellar clouds. The circumstellar envelopes of carbon-rich stars are the heart of the most complex carbon chemistry, which is analogous to soot formation in candle flames or industrial smoke stacks (Henning and Mutschke, 2004). It is expected that carbon grains forming in the outflows of carbon-rich Asymptotic Giant Branch stars are composed of a wide variety of structures (Henning and Salama, 1998). Those stars are in a late stage of their evolution and lose mass via a dense and dusty outflow.

Molecular synthesis occurs in the circumstellar environment on timescales as short as several hundred years (Kwok, 2004). Acetylene (C_2H_2) appears to be the precursor for the synthesis of hexagonal aromatic rings of carbon atoms. Benzene detection has been claimed in the Infrared Space Observatory spectrum of the circumstellar envelope CRL 618 (Cernicharo *et al*., 2001). Recent laboratory data have shown that benzene could be available for aromatic chemistry when sufficiently shielded in circumstellar envelopes from protons and UV photons (Ruiterkamp *et al*., 2005a). Benzene chemistry is the first step in the formation of polycyclic aromatic hydrocarbons (PAHs), fullerene-type material, and larger aromatic macromolecular material (Frenklach and Feigelson, 1989; Cherchneff *et al*., 1992). Those materials make up the major fraction of carbon in space.

FROM THE INTERSTELLAR MEDIUM TO THE TERRESTRIAL PLANETS

Currently 143 gas-phase molecules, many of them organic, have been identified in interstellar space, and ~60 molecules have been detected in circumstellar environments (see http://www.astrochemistry.net) (Charnley *et al*., 2003). Many of the Si-bearing and metal-bearing molecules are only found in circumstellar shells. Apart from silicates, many constituents have been proposed to be present in interstellar grain mantles, including amorphous carbon, hydrogenated amorphous carbon, diamond, refractory organics, and carbonaceous networks such as coal, soot, graphite, quenched-carbonaceous condensates, and others (for reviews, see Henning and Salama, 1998; Ehrenfreund and Charnley, 2000).

Interstellar carbon compounds and their abundance

PAHs are observed in galactic and extragalactic regions, and represent the most abundant carbonaceous gas-phase molecules in space (Tielens *et al*., 1999; Yan *et al*., 2005). Laboratory studies and theoretical calculations have provided important insights into their size and charge state distribution (*e.g*., Salama *et al*., 1996; Allamandola *et al*., 1999; Ruiterkamp *et al*., 2005b). All three isoforms of carbon—diamond, graphite, and fullerene—have been identified in space environments (Cataldo, 2004). Diamonds were proposed to be the carriers of the 3.4 and 3.5 μm emission bands (Guillois *et al*., 1999) observed in planetary nebulae. Graphite has not been unambiguously identified in the interstellar medium but is present in low abundances in meteorites (Nuth, 1985). The third isoform of carbon is the polyhedral C_{60} fullerene first discussed by Kroto *et al*. (1985). It has been suggested that fullerenes may be formed in small amounts in envelopes of R Coronae Borealis stars (Goeres and Sedlmayr, 1992). Rietmeijer *et al*. (2004) showed that cosmic soot analogs consist of close packed metastable C_{60} and giant fullerenes. Fingerprints of the C_{60}^+ ion were discovered in the near-infrared spectra of stars crossing the material of diffuse interstellar clouds (Foing and Ehrenfreund, 1994, 1997). Higher fullerenes have been identified in meteorites (Becker and Bunch, 1997). Laboratory simulations in combination with interstellar observations support the idea that the predominant fraction of carbon in space is present as solid macromolecular carbon (*e.g*., Pendleton and Allamandola, 2002) or amorphous and hydrogenated amorphous carbon (Mennella *et al*., 1998; Duley and Lazarev, 2004; Dartois *et al*., 2005). Fullerenes (Iglesias-Groth, 2004) or defective carbon "onions" (Tomita *et al*., 2004) have also been proposed.

Cosmic abundances in the interstellar medium are derived by measuring elemental abundances in stellar photospheres, the atmospheric layer just above the stellar surface. Such measurements indicate the amount of elements available for the formation of molecules and particles. Cosmic dust models indicate that up to 80% of the carbon in the photon-dominated diffuse interstellar medium is incorporated into solid aromatic macromolecules and gaseous PAHs (Mennella *et al*., 1998; Tielens *et al*., 1999). CO gas and C-based ice species (such as CO, CO_2, CH_3OH, and others) may make up to ~25% of the carbon in cold dense interstellar regions. It must be taken into account that elemental abundances (*e.g*., of carbon) in stars and the interstellar medium may not be the same. Stellar photospheres may not completely represent the composition of interstellar material from which stars are formed, resulting either from the possible underestimation of the degree of heavy element settling in the stellar atmosphere or from incomplete incorporation of heavy elements in stars during star formation (Li, 2005). However, prebiotic molecules (as used in life as we know it) observed in the interstellar

medium have negligible abundances compared with aromatic moieties; organic material in space is predominantly macromolecular and aromatic in nature.

Formation of a solar system and extraterrestrial delivery

Interstellar gas and icy or refractory dust (composed of silicates and carbonaceous material) provide the raw material for the formation of stars and planets. Gravitational collapse of interstellar clouds leads to the formation of a protoplanetary region with a central condensation developing into a star surrounded by a disk (Mannings *et al*., 2000; Boss, 2004). During the formation of a solar system, interstellar gas and dust are mixed, processed, and partly destroyed according to the distance from the forming star (Chick and Cassen, 1997). Radiation chemistry involving x-rays and UV light may also act on close or upper disk layers. Turbulent motions lead to radial mixing of the products within the disk (Irvine *et al*., 2000; Ehrenfreund *et al*., 2004; Markwick and Charnley, 2004). Within disks, dust assembles into grains, which in turn assemble into boloids and ultimately into kilometer-sized planetesimals that interact gravitationally (Weidenschilling and Cuzzi, 1993). The key steps in this assembly process are still a mystery, but once self-gravitation begins planetesimals grow by collision to form terrestrial planets (Blum, 2004).

Many planetesimals formed in such a manner could not be integrated into planets, and their ultimate fate was dependent on their location within the nascent solar system. The giant planets played an important role in stabilizing the structure of the solar system by trapping remnant bodies within defined regions such as the asteroid belt, or by ejecting them into outer solar system regions (Kuiper Belt) and beyond (Oort Cloud). The deflection of small bodies in the inner solar system led to large impacts on the forming terrestrial planets. Comets probably contributed most of the carbonaceous compounds during the heavy bombardment phase 4.5–4 billion years ago (Ehrenfreund *et al*., 2002). Fragments of asteroids and comets such as interplanetary dust particles and carbonaceous meteorites were probably among the other major extraterrestrial contributors of carbon (Chyba *et al*., 1990; Chyba and Sagan, 1992; Oró and Lazcano, 1997).

Comets are principally formed in the region beyond Jupiter. They are predominantly icy bodies that contain some silicates and refractory organic material (Greenberg, 1998). More than 50 molecules have been identified in cometary comae (Crovisier, 2004). Many small organic molecules observed in cometary comae probably originate wholly or partially from the decomposition of larger molecules or particles, which indicates that large polymers such as polyoxymethylene and HCN-polymers may be present in comets (see Ehrenfreund *et al*., 2004). Carbonaceous meteorites contain a substantial amount of carbon (up to 3% by weight) and exhibit evidence of thermal and aqueous alteration believed to have occurred on their parent bodies. Up to 90% of this carbon is macromolecular (Gardinier *et al*., 2000; Sephton *et al*., 2000; Cody and Alexander, 2005). The insoluble macromolecular matter in carbonaceous meteorites is composed of mono- and polyaromatic units bonded with oxygen, sulfur, and small aliphatic chains. More than 70 amino acids have been identified in the soluble fraction of such samples in addition to many other organic compounds, including N-heterocycles, carboxylic acids, sulfonic and phosphonic acids, and aliphatic and aromatic hydrocarbons (Cooper *et al*., 1992; Botta and Bada, 2002; Sephton, 2002). The presence of organic matter in both anhydrous and hydrated interplanetary dust particles has been discussed (Clemett *et al*., 1993; Flynn *et al*., 2003).

What material was delivered to the early Earth, and what are the implications for life?

The large quantities of extraterrestrial material delivered to young terrestrial planetary surfaces in the early history of our solar system may have provided the material necessary for the emergence of life. Tables 2 and 3 summarize the organic material that has been measured to date in comets and meteorites. Data are predominantly compiled from the analysis of the Murchison meteorite and observations of bright comets (*e.g*., for a review see Ehrenfreund *et al*., 2002; Crovisier, 2004). Recent bright comets such as Hale-Bopp have been investigated by spectroscopy from the UV to the radio range in unprecedented detail. Though several classes of organic compounds important in contemporary biochemistry appear on the list of meteoritic compounds, the dominant form of carbonaceous material is aromatic. This is similar to the original interstellar cloud material as discussed in The Formation and Evolution of Organic Material in Space and From the Inter-

Table 2. Production Rates Relative to Water of Organic Molecules in the Coma of Comet C/1995 O1 Hale-Bopp

Molecule	*C/1995 O1 Hale-Bopp*
H_2O	100
CO	12–23
CO_2	6
CH_4	1.5
C_2H_2	0.1–0.3
C_2H_6	0.6
CH_3OH	2.4
H_2CO	1.1
HCOOH	0.09
$HCOOCH_3$	0.08
CH_3CHO	0.02
NH_2CHO	0.015
HCN	0.25
HNCO	0.10
HNC	0.04
CH_3CN	0.02
HC_3N	0.02
OCS	0.4
CS_2	0.2
H_2CS	0.05

Data are taken from Bockelée-Morvan *et al*. (2004).

stellar Medium to the Terrestrial Planets. This aromatic material has not been considered useful for biogenesis; however, aromatic macromolecular materials are known to fragment upon heating, oxidation, or hydrolysis into subunits that may have prebiotic roles. The large quantities of aromatic material compared with trace concentrations of amino acids, N-heterocycles, and sugars arriving *via* extraterrestrial delivery suggest that it might be useful to consider how aromatic material could have contributed to life's origin. The resistance of aromatic compounds to thermal and radiation processing in space is reflected in their high abundance. Consequently, this may apply to the young Earth as well.

TERRESTRIAL SOURCES OF ORGANIC MATTER AND POSSIBLE PREBIOTIC SCENARIOS

In addition to extraterrestrial sources, endogenous synthesis on Earth from atmospheric reactions, hydrothermal vents, or serpentinization may have contributed to the pool of precursor molecules and complex organics. Atmospheric synthesis of organics has been a favorite scenario for many decades. The primitive atmosphere may, however, have been a minor contributor to the organic inventory on the young Earth if it was not reducing. The limited knowledge of the exact density and composition of the atmosphere and the temperature and radiation conditions on the Earth makes it difficult to estimate the relative contributions of extraterrestrial delivery and endogenous atmospheric synthesis. During heavy impacts in the first 500 million years, the Earth's climate may have alternated between exceedingly hot periods (in which most organic compounds were destroyed) and cold periods, making it difficult for life to become established. The concentration of organic material in the prebiotic environment is of crucial importance. Precursor molecules need to be concentrated to react before being degraded by the environment. Small molecules, crucial for living cells in modern chemistry, such as amino acids, heterocycles, and sugars, are rather fragile compounds with regard to temperature, extremes of pH, and radiation.

Table 3. Abundances and Type of Organic Matter in the Best-Studied Carbonaceous Meteorite, Murchison (CM2)

	Abundances	
Compound class	*%*	*ppm*
Macromolecular material	1.45	
CO_2		106
CO		0.06
CH_4		0.14
Amino acids		60
Aliphatic hydrocarbons		12–35
Aromatic hydrocarbons		15–28
Fullerenes		>1
Carboxylic acids		332
α-Hydroxycarboxylic acids		15
Dicarboxylic acids		26
Pyridinecarboxylic acids		>7
Basic N-heterocycles		0.05–0.5
Pyrimidines (uracil and thymine)		0.06
Purines		1.2
Benzothiophenes		0.3
Dicarboximides		>50
Amines		8
Amides		
Linear		>70
Cyclic		>2
Alcohols		11
Aldehydes		11
Ketones		16
Sugar-related compounds		~24
Urea		25

Data are compiled from Botta and Bada (2002), Sephton (2002), and Sephton and Botta (2005).

Therefore, depending on the rates of synthesis, it may be questionable as to whether such compounds could have been involved in the formation of the first living entities.

Early Earth conditions

The three main requirements for life as we know it are organic compounds, liquid water, and free energy. These may also have been crucial for the establishment of the first living systems. It seems reasonable to assume that life required a solvent for reactions to occur in (water), compounds for the living system to be built from (most likely organic compounds), and free energy to allow for chemical evolution to proceed both by providing the driving force for chemical reactions and for allowing fluctuations in the environment that would provide both energy and a natural selection pressure for the evolution of the system. The early solar system appears to have provided all of these, and it seems possible that these conditions could be present on numerous other bodies in the galaxy. The synthesis of small organic compounds from simple gas mixtures is one of the best experimentally understood steps in the origin of life. The efficiency of organic production and the nature of the products depend on the specific type of energy introduced and the gas mixture used.

There is little agreement on the composition of the primitive atmosphere. Opinion varies from strongly reducing ($CH_4 + N_2 + NH_3 + H_2O$, or $CO_2 + H_2 + H_2O + N_2$) to neutral ($CO_2 + N_2 + H_2O$), though it is generally believed that free O_2 was absent (Canuto *et al*., 1983). There has been less experimental work with gas mixtures that contain CO or CO_2 as carbon source in place of CH_4, though CO-dominated atmospheres could not have existed except transiently (Miyakawa *et al*., 2002a). As mentioned earlier, the efficiency of organic production depends on the reducing nature of the atmosphere; neutral atmospheres may be some 10^5-fold less efficient at producing organics than reducing atmospheres (Miller, 1998). Recently, the possibility that the early atmosphere contained significant amounts of H_2 has been resuscitated (Tian *et al*., 2005).

The early Earth can be viewed as a dynamic system warmed by heat from accretion, the sinking of iron into the core, impacts, and decay of radioactive elements. The most abundant energy sources on Earth today are light, electric discharges from lightning and static electricity, radioactive decay, and volcanism (Table 4). Energy fluxes from these sources may have been considerably different in the primitive environment. For example, the primitive Sun would have provided a much higher flux of UV radiation, though the total luminosity was lower. It is likely that volcanic activity was more intense and energy from radioactive decay was more abundant (Mosqueira *et al*., 1996). Shock waves from extraterrestrial impactors and thunder were also probably more common during the planetary accretion process. It is difficult to es-

Table 4. Fluxes of Energy from Various Sources on the Present Earth: Solar Fluxes at Sea Level

Source	*Energy (cal/cm²/year)*	*Energy (J/cm²/year)*
Total radiation from Sun	260,000	1,090,000
UV light		
<300 nm	3,400	14,000
<250 nm	563	2360
<200 nm	41	170
<150 nm	1.7	7
Electric discharges	4.0[a]	17
Cosmic rays	0.0015	0.006
Radioactivity (to 1.0 km)	0.8	3.0
Volcanoes	0.13	0.5
Shock waves	1.1[b]	4.6

Fluxes on the primitive Earth may have differed, in particular, the flux of shorter wavelength UV light, especially if there was increased UV output by the early Sun, and in the absence of ozone and molecular oxygen in the primitive atmosphere. The majority of the solar flux incident on the upper atmosphere is presently in the form of longer wavelength radiation ($\lambda > 300$ nm) of lower energy per photon, which is unable to engage in useful atmospheric prebiotic chemistry. Adapted from Miller and Orgel (1974).

[a]3 cal/cm²/year of corona discharge and 1 cal/cm²/year of lightning.

[b]1 cal/cm²/year of this is the shock wave of lightning bolts and is also included under electric discharges.

timate the difference in electric discharge flux early in the Earth's history.

The Sun is expected to have followed a typical stellar evolution for its mass and spectral type (G2 class). It is expected that solar luminosity would have been ~30% less around the time of the origin of life (Kasting and Catling, 2003). A possible consequence of this is that the prebiotic Earth may have frozen completely to a depth of ~300 m (Bada *et al*., 1994), though there is evidence that liquid water was present ~4 billion years ago (Mojzsis *et al*., 2001; Wilde *et al*., 2001). The presence of liquid surface water would have required that the early Earth maintained a heat balance that offset the lower solar flux from the faint young Sun. Atmospheric greenhouse warming from CH_4, NH_3, or CO_2 has been proposed, but this is still debated (Kasting and Catling, 2003).

Prebiotic synthesis of amino acids, nucleobases, and sugars

The first successful amino acid synthesis under prebiotic conditions was carried out with an electric discharge and a strongly reducing model atmosphere of CH_4, NH_3, H_2O, and H_2 (Miller, 1953). This experiment produced a large yield of racemic amino acids, together with hydroxy acids, short aliphatic acids, and urea. The products obtained give clues to the mechanism of synthesis. The yield of both α-amino and α-hydroxy acids of the same carbon skeletons suggested that the compounds were formed by the Strecker amino acid and cyanohydrin hydroxy acid syntheses. These reactions occur when aldehydes or ketones, ammonia, and cyanide (derived from gas-phase reactions of CH_4, NH_3, H_2O, and H_2) are allowed to react in water. Detailed studies of the equilibrium and rate constants of these reactions have been performed (Miller, 1957). The results demonstrate that both amino and hydroxy acids can be synthesized at high dilutions of HCN and aldehydes in a simulated primitive ocean. The reaction rates depend on temperature, pH, and reactant concentrations, but are rapid on a geologic time scale. It was determined that such reactions could occur starting from even extremely low concentrations of precursors, as might be delivered to the oceans and even from relatively low-yielding atmospheric synthesis in neutral CO_2/N_2 atmospheres (Schlesinger and Miller, 1973, 1983).

The Strecker synthesis of amino acids requires the presence of NH_3 in the prebiotic environment. Gaseous NH_3 is rapidly decomposed by ultraviolet light (Kuhn and Atreya, 1979), and during early Archean times the absence of a significant ozone layer would have imposed an upper limit to its atmospheric concentration. Since NH_3 is soluble in water, most of the Earth's NH_3, if the primitive oceans and sediments were buffered to the modern value of pH ~8, would have been present as dissolved NH_4^+ (the pK_a of NH_3 is ~9.2) in equilibrium with dissolved NH_3.

Nucleic acids are the central repository of the information organisms use to construct enzymes *via* protein synthesis. One of the principal characteristics of life is the ability to transfer information from one generation to the next. Nucleic acids seem uniquely suited for this function, and thus a considerable amount of attention has been dedicated to elucidate their prebiotic synthesis. The first evidence that the components of nucleic acids could have been synthesized non-biologically was provided in 1960 when it was found that concentrated solutions of ammonium cyanide refluxed for a few days produced adenine (Oró, 1960; Oró and Kimball, 1961). Other purines, including guanine, hypoxanthine, xanthine, and diaminopurine, have since been produced using variations of this synthesis (Sanchez *et al*., 1968). The kinetics of the intermediate reactions in this synthesis have been used to delineate the limits of geochemically plausible synthesis. The steady-state concentrations of HCN would have depended on the pH, the temperature of the early oceans, and the input rate of HCN from atmospheric synthesis (itself depending on the reducing nature of the atmosphere) or extraterrestrial delivery. Assuming favorable production rates, steady-state concentrations of HCN of 2×10^{-6} *M* at pH 8 and 0°C in the primitive oceans were estimated (Miyakawa *et al*., 2002b). At 100°C and pH ~8 the steady-state concentration was estimated as 7×10^{-13} *M*. Oligomerization and hydrolysis compete at approximately 10^{-2} *M* concentrations of HCN at pH 9 (Sanchez *et al*., 1966a), though it has been shown that adenine is still produced from solutions as dilute as 10^{-3} *M* (Miyakawa *et al*., 2002c). Clearly, some concentration mechanism would have been necessary, with eutectic freezing seemingly the most likely, as HCN cannot be concentrated by evaporation (Sanchez *et al*., 1966a).

The prebiotic synthesis of pyrimidines has also been investigated. Cytosine and uracil are synthesized from the reaction of aqueous cyanoacetylene and cyanate (Sanchez *et al*., 1966b), both

of which are produced in electric discharge reactions. Later it was found that cyanoacetaldehyde (produced from cyanoacetylene) and urea react to form cytosine and uracil in high yield when dried together (Robertson and Miller, 1995). The reaction of uracil with formaldehyde and formate gives thymine in good yield (Choughuley *et al*., 1977).

Most biological sugars have the empirical formula $(CH_2O)_n$, a point underscored by Butlerov's (1861) discovery of the formose reaction, which showed that a complex mixture of biologically important sugars could be formed by the reaction of HCHO under basic conditions. The mechanism of the Butlerov synthesis is complex and incompletely understood. It depends on the presence of suitable inorganic catalysts; in the absence of basic catalysts, little or no sugar is obtained. There are three major obstacles to the relevance of the formose reaction as a source of sugars on the primitive Earth. The first problem is that the Butlerov synthesis gives a wide variety of straight-chain and branched sugars. The second problem is that the conditions of synthesis are also conducive to the degradation of sugars (Reid and Orgel, 1967). Sugars undergo various irreversible degradation reactions on geologically short time scales, which would have prohibited their accumulation on the primitive Earth. At pH ~7, the half-life for decomposition of ribose is 73 min at 100°C, and 44 years at 0°C (Larralde *et al*., 1995). The same is true for most other sugars. The third problem is that the concentrations of HCHO required appear to be prebiotically implausible (Pinto *et al*., 1980).

Amphiphilic molecules are especially interesting because of their ability under appropriate conditions of pH, temperature, and concentration to assemble spontaneously into micelles, vesicles, and other types of aggregates (Segré *et al*., 2001). The origin of life picture proposed in the "Lipid World" was originally developed by Luisi *et al*. (1988) and by Morowitz *et al*. (1988), who worked toward assembling protocells based on a self-reproducing lipid vesicle encapsulating a self-replicating RNA protogene. Simple fatty acids are found in meteorites and were probably available at the time of the origin of life. Encapsulation of self-assembling structures, such as microtubules, within liposomes has also served as an incomplete attempt to develop a protocell model (Hotani *et al*., 1992). An interesting twist to the Lipid World came recently with the proposed role of atmospheric aerosols in the origins of life (Tuck, 2002).

Most prebiotic simulations do not generate large amounts of fatty acids, with the exception of certain hydrothermal vent simulations, which arguably use unreasonably high concentrations of reactants (McCollom *et al*., 1999). The Murchison meteorite contains small amounts of higher-molecular-weight straight-chain fatty acids, some of which may be contamination (Yuen and Kvenvolden, 1973). Amphiphilic components have been positively identified in the Murchison meteorite (Deamer, 1985). It should be noted that phospholipids can also be synthesized fairly readily from simple prebiotic starting materials under simulated drying beach conditions (see Segré *et al*., 2001, and references therein).

Energy-transducing molecules have not been as well investigated, if only because "metabolism-first" origins of life theories remain less well elucidated (Smith and Morowitz, 2004). However, a number of potential electron carrier molecules and photon-harvesting molecules have been reported in prebiotic simulations. Notable among these are porphyrin analogues (Hodgson and Ponnamperuma, 1968; Simionescu *et al*., 1980; Chadha and Choughuley, 1984) and PAHs (Deamer, 1992). As yet, these have not been demonstrated to be useful prebiotically; however, several schemes have been suggested (Stilwell, 1977; Stephan, 2002).

Informational polymers

One popular theory for the origin of life proposes the existence of an RNA World, a time when RNA molecules played the role of both catalysts and genetic molecule in biochemistry (Gesteland *et al*., 1999; Joyce, 2002). A great deal of research has been carried out on the prebiotic synthesis of nucleosides and nucleotides. While considerable progress has been made in this area, a number of chemically implausible steps in this scheme still render it problematic. This raises the intriguing possibility that the first informational polymer may not have been RNA-based, which will be explored in more detail herein. The most promising nucleoside syntheses start from purines and pure D-ribose in drying reactions, which simulate conditions that might occur in an evaporating basin (Fuller *et al*., 1972). Using hypoxanthine and a mixture of salts reminiscent of those found in seawater, up to 8% of α-D-inosine

is formed, along with the β-isomer. Adenine and guanine give lower yields, and in both cases a mixture of α- and β-isomers are obtained (Fuller *et al*., 1972). Pyrimidine nucleosides have proven to be much more difficult to synthesize. Direct heating of ribose and uracil or cytosine has thus far failed to produce uridine or cytidine. Pyrimidine nucleosides have, however, been produced by more complicated multistep schemes involving photochemical rearrangement (Sanchez and Orgel, 1970; Ingar *et al*., 2003).

The difficulties with prebiotic ribose synthesis and nucleoside formation have led some to speculate that perhaps a simpler genetic molecule with a more robust prebiotic synthesis preceded RNA. A number of alternatives have been investigated (Fig. 2). Some substitute other sugars for ribose (Beier *et al*., 1999; Eschenmoser, 2004), including the structurally very simple threose nucleic acid (TNA). These molecules would likely suffer from the same drawbacks as RNA, which include the difficulty of selective sugar synthesis, sugar instability, and the difficulty of nucleoside formation. More exotic nucleoside alternatives have been proposed based on the peptide nucleic acid (PNA) analogues (Nielsen *et al*., 1991) and other acyclic monomers (Joyce *et al*., 1987). Miller and co-workers (Nelson *et al*., 2000) were able to demonstrate the synthesis of the components of PNA under the same conditions required for the synthesis of the biological purines and pyrimidines. The assembly of the molecules into oligomers has not yet been demonstrated and may be unlikely (Eriksson *et al*., 1998). There may be alternative structures that have not yet been investigated but may sidestep some of the problems with the PNA backbone. As of this writing, the simplest acyclic nucleoside analogue, glycol nucleic acid (GNA), has been shown to form stable Watson–Crick base-paired structures (Zhang *et al*., 2005) (see Fig. 2 for examples of some proposed alternative nucleic acid structures).

Phosphates

Condensed phosphates, principally in the form of ATP, are the universal biological energy currency. However, abiological dehydration reactions are extremely difficult in aqueous solution because of high water activity, and it has been suggested that condensed phosphates are unlikely prebiotic compounds (Keefe and Miller, 1995). There is the intriguing possibility, however, that phosphates may have entered early biochemistry in the form of phosphonic acids (De Graaf and Schwartz, 2005), which are present in the Murchison meteorite (Cooper *et al*., 1992). There is some evidence that condensed phosphates are emitted in volcanic fumaroles (Yamagata *et al*., 1991). The heating of ammonium phosphates with urea leads to a mixture of high-molecular-weight polyphosphates (Osterberg and Orgel, 1972). Although polyphosphates are not especially good phosphorylating reagents under prebiotic conditions, they tend to degrade, especially in the presence of divalent cations at high temperatures, to cyclic phosphates such as trimetaphosphate. Trimetaphosphate has been shown to be a phosphorylating agent for various prebiological molecules, including amino acids and nucleosides (Schwartz, 1969; Rabinowitz and Hampai, 1978). Another intriguing possibility is

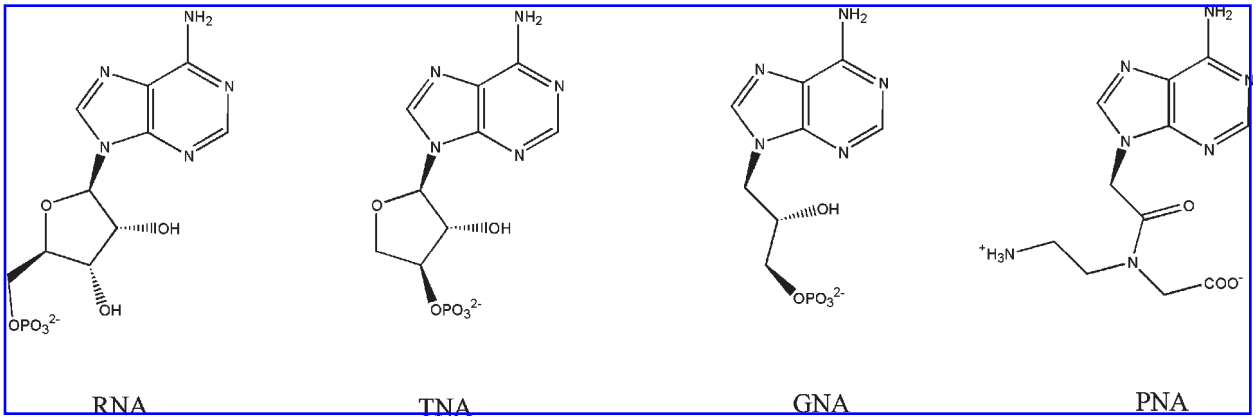


FIG. 2. Some possible monomers that could have been used to construct a primordial genetic molecule.

that reactive phosphates were delivered extraterrestrially in the form of phosphide minerals (Pasek and Lauretta, 2005).

Nucleosides can also be phosphorylated with acidic phosphates such as NaH_2PO_4 when dry-heated (Beck *et al*., 1967) and by heating with ammonium phosphate and urea at moderate temperatures, as might occur in a drying basin (Lohrmann and Orgel, 1971). Yields as high as 20% of mixed nucleotide monophosphates were achieved by heating nucleosides with the mineral hydroxyapatite, urea, and ammonium phosphate (Lohrmann and Orgel, 1971). Interestingly, nucleoside 2′,3′-cyclic phosphates are major products of such syntheses. These have been shown to be polymerizable into oligonucleotides under drying conditions (Verlander *et al*., 1973).

Subsurface synthesis

The discovery of hydrothermal vents at the oceanic ridge crests was a major oceanographic discovery (Corliss *et al*., 1979). A hot microbial biosphere associated with active volcanism along the global mid-ocean ridge network has improved our knowledge of life adapting to extreme environments (Kelley *et al*., 2002). Since hydrothermal circulation probably began early in the Earth's history, it is likely that vents were present in the Archean oceans. It has been suggested that organic compounds are produced during passage through the temperature gradient of the 350°C vent waters to the 0°C ambient ocean waters under high pressure (Corliss *et al*., 1981). Polymerization of the organic compounds thus formed, followed by their self-organization, has also been proposed to take place in these environments. At first glance, submarine hydrothermal springs appear to be ideally suited for creating life, given the geological plausibility of a hot early Earth. Unfortunately, it is difficult to corroborate these ideas with the composition of modern vent effluent, as most of the organic material released from modern sources is degraded biological material, and it is difficult to separate the biotic from the abiotic (nonbiological) components of these reactions (Miller and Bada, 1988; Ferris, 1992; Shock and Schulte, 1998; Holm and Charlou, 2001).

Iron sulfides have been proposed to play a potential catalytic role in the formation of organometallic compounds (Cody *et al*., 2000) and in the origin of life (Russell and Hall, 1997; Russell and Martin, 2004). In the latter case, though, hydrothermal solutions are buffered at lower temperatures (<115°C) and higher pH (10–11). The most elaborate articulation of hydrothermal vent biopoiesis stems from the work of Wächtershäuser (1988), who has argued that life began with the appearance of an autocatalytic, two-dimensional chemolithtrophic metabolic system based on molecular hydrogen and a reduction potential provided by the formation of the highly insoluble mineral pyrite (FeS_2):

$$FeS + H_2S \rightarrow FeS_2 + H_2 \ (E^\circ = 620 \text{ mV}, \Delta G^\circ = -9.23 \text{ kcal/mol})$$

The FeS/H_2S combination is a strong reductant for a variety of organic compounds under mild conditions. The FeS/H_2S system has not been shown to reduce CO_2 to amino acids, purines, or pyrimidines in the laboratory, though there is more than adequate free energy to do so (Keefe *et al*., 1995). However, pyrite formation can produce molecular hydrogen and reduce nitrate to ammonia, and acetylene to ethylene (Maden, 1995). Although speculative, it is possible that under certain as yet unspecified geological conditions the FeS/H_2S combination could have reduced CO and CO_2 released from deep-sea vents, which would have led to biochemical monomers such as amino acids and nucleosides (Orgel, 1998). Peptide synthesis could have taken place in an iron and nickel sulfide system (Huber and Wächtershäuser, 1998) involving amino acids formed by electric discharges *via* a Strecker-type synthesis, though this scenario might require the transportation of compounds formed in other environments to the deep-sea vents (Rode, 1999). Environmental concentrations of reactants may, however, be prohibitively low for such polymerization reactions.

Another reaction pathway that is sometimes discussed in the context of hydrothermal vents is the Fischer–Tropsch type (FTT) synthesis (Ferris, 1992). This reaction has also been proposed as a synthetic mechanism in volcanoes and meteorites as well as on the surface of extraterrestrial mineral grains that enter the Earth's atmosphere (Hayatsu *et al*., 1971, 1972; Basiuk and Navarro-Gonzalez, 1996; Hill and Nuth, 2003). The reaction involves the passage of gases over hot iron, nickel, or silicate catalysts, whereby small organic molecules are generated via surface catalysis. Although the reaction produces many important

biogenic molecules, the reaction mechanisms are poorly understood. While FTT synthesis is an important industrial process, its prebiotic relevance is unclear. The catalysts used in industry are easily poisoned by sulfide, which is a common component of hydrothermal fluid. The concentrations of gases encountered in vent fluids and the high water activities encountered there may render FTT synthesis difficult. To date, it remains controversial as to whether organic molecules are generated abiotically from oxidized carbon in vent environments.

Serpentinization of oceanic crust is another mechanism by which FTT syntheses are proposed to occur (Holm and Andersson, 1998; Sleep *et al*., 2004). Conversion of ultramafic peridotite (olivine and pyroxene) to serpentine and magnetite leads to the formation of H_2, which can reduce CO_2 to CH_4. Magnetite has been shown to convert CO_2/H_2 to organic compounds under conditions reminiscent of serpentinization (Berndt *et al*., 1996). Although serpentinization may have resulted in the formation of hydrocarbons in some special subsurface environments, the temperature conditions under which these syntheses occur would not allow the survival of most biogenic compounds. Serpentinization may, however, also have served as a source of CH_4 and H_2 for atmospheric reactions.

If mineral assemblages in the cooler zones of hydrothermal vent environments were sufficiently reducing, then similar reactions in vent systems may have been a source of atmospheric H_2, CH_4, and NH_3 (Shock *et al*., 1998; Kelley *et al*., 2002, 2005).

In general, organic compounds are more readily decomposed than created at hydrothermal vent temperatures, though this would not apply to the low temperature off-ridge vents (*e.g*., Russell and Martin, 2004). Most biological molecules have half-lives to hydrolysis on the order of minutes to seconds at the high temperatures associated with hydrothermal vents (Table 5). As noted earlier, ribose and other sugars are extremely thermolabile (Larralde *et al*., 1995). Pyrimidines, purines, amino acids, nucleotides, and peptides are nearly as labile (White, 1984; Shapiro, 1995). The half-lives for polymers are even shorter as there are many potential breakage points. However, it is possible that hydrothermal vents may serve as synthesis sites for simpler compounds such as acetate (Russell and Martin, 2004) or more complex organic compounds such as fatty acids (Shock *et al*., 1998). The adsorption equilibria of most organics with mineral surfaces tend to be higher at lower temperatures; thus adsorption on mineral surfaces would tend to concentrate any organics created at hydrothermal vents in cooler zones (Sowerby *et al*., 2001).

Table 5. Half-Lives for the Decomposition of Some Important Biomolecules at pH 7 and Various Temperatures

	$t_{1/2}$ *(years) at*	
Compound	*0°C*	*100°C*
Amino acids	10^6–10^{10}	1–10^4
Sugars (ribose)[b]	44	73 min
Nucleobases (ACGTU)[c]	10^4–10^9	0.05–56
Nucleosides (uridine)[d]	10^6	400
RNA scission[d]	900	0.01
DNA depurination[d]	10^5	0.06
Peptide hydrolysis[d,e]	10^8	~1

References: [a]Miller and Orgel (1974); [b]Larralde *et al*. (1995); [c]Levy and Miller (1998); [d]White (1984); [e]Snider and Wolfenden (2000).

The concentrations of biomolecules that could have accumulated on the primitive Earth are governed largely by the rates of production and destruction. Although it is presently not possible to state which compounds were essential for the origin of life, it does seem possible to preclude high temperature environments even if simple organic compounds used in modern biochemistry were involved (Cleaves and Chalmers, 2004).

Much of the upper layers of the Earth's crust are porous. Rock and sediment contain pore spaces between grains, within fractures, or in cavities (Gold, 1999). Both extraterrestrially and atmospherically delivered organic compounds, and in particular subsurface-synthesized organic compounds, might be expected to exist in the pore space within the upper layers of the planetary crust. Water, organic compounds, richly varied catalytic minerals, and free chemical redox energy are all expected to be found within pore spaces. The amount of pore space on a young planet would have been an immense volume and has been proposed as a candidate site for the origin of life (Colgate *et al*., 2003).

Although significant advances have occurred as a result of more than 50 years of research in prebiotic chemistry using compounds crucial in modern biochemistry, there have been no major breakthroughs in the understanding of the origin of life. Furthermore, early Earth conditions may

have been extremely variable and hostile. This raises the questions of whether fragile species such as amino acids, nucleobases, and sugars would have been abundant enough to have been useful for the origin of life.

FUNCTIONAL ASSEMBLIES AND THEIR INTEGRATION INTO A MINIMAL LIFE FORM

Primitive life could, in principle, be based on a radically different chemistry and organizational structure than those of contemporary life. Such alternative scenarios are not discussed here, though they are actively pursued in the artificial life research communities both in simulations and as the foundation for robotics design (Brooks, 2001). In the context of the origins of life, it is not obvious which alternative chemistries or organizational principles should be investigated. It is possible that life based on alternative chemical and organizational principles evolved into life as we know it. For example, clay and other inorganic and organic compounds could have played a key role both as precursors and as catalysts (Cairns-Smith, 2005).

As discussed in From the Interstellar Medium to the Terrestrial Planets and in Terrestrial Sources of Organic Matter and Possible Prebiotic Scenarios, small molecules, such as amino acids and N-heterocycles, are easily degraded by temperatures much above 50°C and UV radiation (Tables 5 and 6), whereas larger aromatic structures can resist far higher doses of both. Amino acids have very short half-lives (Peeters *et al*., 2003), and N-heterocycles are more easily destroyed than their carbonaceous cognate molecules such as benzene (Table 6). Heterocycles that contain several N-atoms in the ring, such as adenine, have decreased half-lives when exposed to UV radiation (Peeters *et al*., 2003, 2005).

As stated earlier, the abundant carbonaceous macromolecular material predominant in space environments consists mainly of aromatic units connected by aliphatic bridges. Benzene (C_6H_6), the most simple aromatic hydrocarbon, is unreactive to common double bond transformation and, when forced to react (by higher temperature and/or a catalyst), undergoes substitution reactions rather then addition reactions. The enhanced stability (also called aromatic stabilization) of benzene is 36 kcal/mol. PAHs are fused benzene rings, and in particular large catacondensed PAHs (with decreased H/C ratio) are unusually stable to chemical modification. Aromatic units would therefore not, in particular, be rapidly destroyed by temperature and radiation when incorporated in a three-dimensional macromolecule. The structural similarity among the aromatic fractions of the insoluble matter of carbonaceous meteorites indicates that these aromatics were incorporated at an early stage and remained stable despite low temperature hydrothermal reactions (Cody and Alexander, 2005). Chemical oxidation of such macromolecules may yield soluble organic acids. Laboratory studies investigating kerogen-like material confirm an increase in aromaticity upon pyrolysis under vacuum (Ehrenfreund *et al*., 1991). Infrared spectroscopy has shown that, upon processing kerogens, the more fragile aliphatic bonds decrease and the aromatic (C-C) ring bonds become predominant.

Functional requirements for early life

It is assumed that life was initially simpler than, yet functionally similar to, life as we know it today.

Table 6. UV Stability of Heterocycles Measured in the Laboratory in Ar Matrix

	Half-life			
Species	*Lab (s)*	*DISM (years)*	*DC ($\times 10^6$ years)*	*1 AU (min)*
Benzene	220	32	3.2	54
Pyridine	123	18	1.8	32
Pyrimidine	56	8	0.8	14
s-Triazine	17	2.5	0.2	4

Half-lives were calculated from the laboratory data of Peeters *et al*. (2005) and extrapolated to diffuse interstellar environment (DISM), dense clouds (DC), and the solar system at 1 AU from the Sun using the appropriate UV fluxes: 1×10^8 photons/cm^2/s for the DISM (Mathis *et al*., 1983), 1×10^3 photons/cm^2/s for DC (Prasad and Tarafdar, 1983), and 3×10^{13} photons/cm^2/s for the solar UV flux >6 eV at 1 AU distance from the Sun.

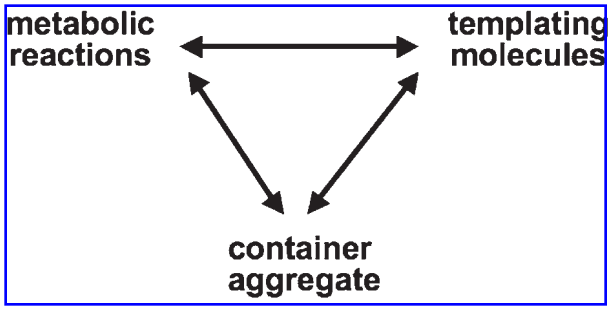


FIG. 3. A simple molecular system could be considered alive if it converts resources into building blocks, grows, replicates, and evolves. Thus a minimal protocell consists of a cooperative structure that combines three functionalities: (1) a metabolism that can convert resources into building blocks, (2) a set of inheritable (templating) molecules that influence the overall growth of the protocell, and (3) a container that keeps (1) and (2) together and acts as a physical structure that facilitates the necessary chemistry for self-reproduction.

A simple molecular system could be considered alive if it turns resources into its own building blocks, grows, replicates, and evolves. This assumes that primitive life would have, at one point, consisted of a compartmentalized genetic system coupled with an energy-harvesting process. While it is difficult to define minimal life, this operational definition is based on three functionally interconnected processes: a metabolic mechanism, a genetic mechanism, and a method of keeping these together through encapsulation (Fig. 3). The metabolic process harnesses energy necessary to convert resource materials into building blocks. The living structure uses the building blocks to grow and replicate. The inheritable molecules influence the growth processes, and as replication proceeds, with some copying error, the functional abilities of the system also change. The accumulated functional differences define the selective advantage of the structures, which enables evolution. The question is, which prebiotically available molecules could have served as container, metabolic, and genetic molecules?

The original question of how prebiotic building blocks can assemble and transform themselves into a minimal living system can now be broken down into two coupled questions:

1. What prebiotic resources can potentially form building blocks, and how can they subsequently be transformed into containers, metabolic reaction networks, and informational polymers?
2. How can these three components assemble into the cooperative organization of a protocell that satisfies the minimal requirements for a living system?

From the astronomical point of view, solid aromatic material is likely to be the most abundant. Aromatic material may be suited to form container compounds, could act as a mediator in metabolic pathways, and could potentially be assembled into informational polymers (Fig. 4 shows typical PAH structures). Table 7 summarizes organic matter identified in the carbona-

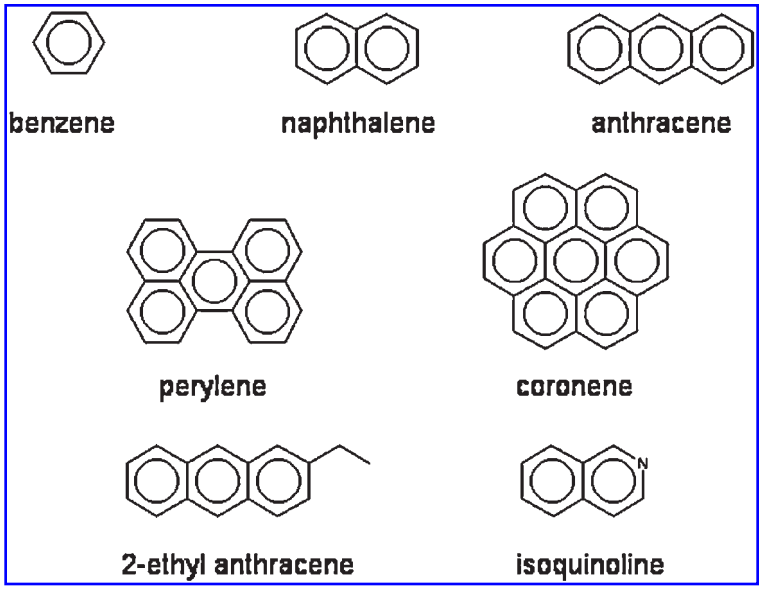


FIG. 4. Examples of typical PAH structures and compounds discussed in Functional Assemblies and Their Integration into a Minimal Life Form.

Table 7. Proposed Chemical Building Blocks Based on the Reported Finding from Table 3

Prebiotic resources	*Building blocks*	*Transformed structures*	*Possible role in a protocell*
Carboxylic acids, amine salts, and charged PAH	Lipid aggregates	Micelles, vesicles, bilayer, etc.	Container
Aliphatic hydrocarbons, carboxylic acids	Block co-polymer	Giant micelles, vesicles	Container
Aromatic hydrocarbons	Polyaromatic compounds	Liquid crystal phase, calixarenes, etc.	Informational polymer, metabolic network components as energy transducer and inceptor
Amides, dicarboxamides, amines, carboxylic acids	Polyamides, PNA, etc.	double strands, self-assembling of peptide	Informational polymer, metabolic network components as energy transducer and inceptor
Sugar-related compounds	Oligosaccharides	Vesicles, cyclodextrins	Container
Fullerenes	Fullerene complexes		Metabolic network components as energy transducer

Each molecular component is investigated for its potential functional properties as container components, genetic elements, or metabolic elements.

ceous meteorite Murchison (Table 3) as potential prebiotic resources, their corresponding building blocks, structures that can be transformed, and their possible roles in a protocell. We will first discuss the details of transformed structures, which can potentially contribute to each function (*i.e.*, container, metabolic compound or informational polymer), and then discuss their possible integration into a living system.

Possible container chemistry

The search for plausible PAH-like molecules that can act as container building blocks is similar, in spirit, to the search for factors that led to the compartmentalization of lipids, a key step in the evolution of life (Luisi *et al*., 1998; Szostak *et al*., 2001; Apel *et al*., 2002; Monnard and Deamer, 2002). Noncovalent protocellular assemblies, generated by catalyzed recruitment of diverse amphiphilic and hydrophobic compounds, could even have constituted the first systems capable of information storage, inheritance, and selection (Segré and Lancet, 2000). Containers can be formed by noncovalent self-assembly. Self-assembly usually requires asymmetric building blocks in the sense that different parts of the building blocks interact differently with the solvent, which is typically water. All bilayer-forming molecules are amphiphiles, with a hydrophilic "head" and a hydrophobic "tail" on the same molecule. Because it is unlikely that complex lipid biosynthesis pathways were present in the earliest forms of life, it seems reasonable that the protocell used lipid-like molecules available in the environment. As noted earlier, PAH derivatives, both free and in the form of kerogen-like polymer, represent over 90% of the organic material of carbonaceous meteorites (see From the Interstellar Medium to the Terrestrial Planets). If there was substantial survival of the organic content of meteoritic and cometary infall during late accretion of the Earth, aromatic molecules would presumably be major components of the organic inventory. Indeed, many PAHs exhibit such structures and are able to self-assemble into bilayer membranes. For example, isoquinoline or naphthalenecarboxylic derivatives are known to form bilayer structures (Chen *et al*., 1999).

Depending on the molecular building blocks, self-assembly is driven by weak forces, such as coulombic attractions, π-π stacking interactions, van der Waals, hydrogen bonding, and hydrophobic forces, or by favorable entropy changes resulting from the release of interfacial water molecules. Small amphiphilic molecules such as long-chain fatty acids, amphiphilic quaternary amines,

and charged PAHs can easily form aggregates in water and consequently assemble into micelles, vesicles, or more complex structures with typical diameters of less than 10 to a few hundred nanometers, depending on conditions (pH, salt concentration, and temperature). Mixed with hydrocarbons or hydrophobic PAHs, the surfactant molecules can encapsulate or integrate more oily substances. In nonpolar media, surfactants can also form inverse micelles around water droplets.

While small lipid molecules of sufficient carbon length may form self-assembling structures, other polymers such as block co-polymers also have a tendency to form stable supramolecular noncovalent assemblies (Discher and Eisenberg, 2002). The building blocks for such polymers (hydrocarbons, carboxylic acids, polyesters, polyamides, oligosaccharides, and polyaromatic compounds) could presumably have been synthesized on early terrestrial planets. As polymeric amphiphilic molecules, these supramolecular noncovalent assemblies can be spontaneously transformed to giant micelles or vesicles with diameters of up to tens of micrometers (Fig. 5).

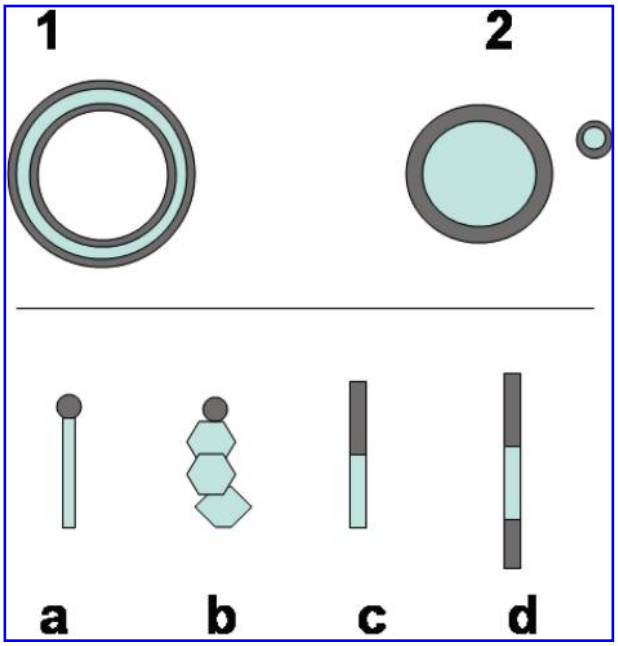


FIG. 5. Possible minimal containers: (1) water-filled bags and (2) hydrophilic bags. In (1), the metabolic and the genetic functionalities reside within the water volume inside the vesicle. In (2), the metabolic and the genetic components are either hydrophobic or amphiphilic and located at the exterior water–lipid interface or within the hydrophobic interior. Note that these container structures can range in size from less than 10 nm to hundreds of micrometers. (**a**) A conventional amphiphilic molecule with a polar head group and a non-polar tail. (**b**) A PAH polymer with a polar head group. (**c** and **d**) Block copolymers with one (c) or two (d) hydrophilic sections. The amphiphilic molecule in (a) and certain (b) and (c) polymers can form (bi- or multilayer) vesicles, while all of them can form micelle-like structures (spheres or rods), depending on pH, salt concentration, and other external factors.

Possible metabolic chemistry

Early terrestrial planets have immense free energy reserves available both as sunlight and as chemical redox energy within the crustal minerals and its supply of organics. Chemolitotrophs (organisms that get their energy from oxidation of inorganic compounds) could have emerged with a metabolism based on the nonequilibrium chemical energy in the subsurface minerals. Such simple chemolitotrophic life forms could have harnessed energy from a variety of chemical species such as S, H_2S, Fe^{3+}, NH_4^+, CO_2, or H_2. Mineral candidates include magnetite (Fe_3O_4) and maghemite (cubic Fe_2O_3), and the iron sulfides pyrite (FeS_2), greigite (Fe_3S_4), and pyrrhotite (Fe_7S_8). However, the first life forms may not have been lithoautotrophs, as the primitive Earth most likely had a host of complex organics available. It seems reasonable to assume that the first life forms would have utilized the simplest possible metabolic processes such as fermentation and not more sophisticated energy-efficient metabolisms. Later, life may have depleted the initially easily available free energy sources, which would have favored the evolution of more efficient metabolisms.

The large inventories of endogenously generated and extraterrestrially delivered organics could have acted as the earliest metabolic energy sources as well as building blocks. Light energy today is captured by photosynthetically sensitizing systems of plants and prokaryotes. What sensitizer molecules may have been available in the prebiotic environment? Again, PAHs and their derivatives may be good candidates (Deamer, 1992). Many PAHs and their derivatives all absorb light in the near-UV and blue region, and could function to capture light energy, either by donating electrons to produce molecules with higher chemical potential or by generation of ionic gradients. Many aromatic hydrocarbons, fullerenes, and PAHs appear to be capable of capturing or transducing energy (see Table 7). For example, PAHs, such as anthracene and quinones (Chen *et al.*, 1998a, 1999), and other heterocyclic aromatic molecules can be excellent charge trans-

fer mediators or photosensitizers that might be able to carry out redox or photochemical reactions such as the conversion of lipid precursors to lipid molecules. Fullerenes and their derivatives have been used widely as light-harvesting and electron transfer components in photovoltaic devices (Taylor and Walton, 1993). Since a metabolism functions by transforming resources into building blocks, slightly modifying one type of organic compound into a usable building block may have been the first step in the development of metabolism (Horowitz, 1945), especially if such processes were thermodynamically favorable (Fig. 6).

Possible template chemistry

A variety of functional amphiphilic containers and energy-capturing and -transduction molecules likely existed on the prebiotic Earth. Templating molecules, however, might be more problematic to synthesize prebiotically. Despite their more complex structure, both simple PAH-based templating polymers and the more sophisticated XNA (*e.g.*, RNA, TNA, GNA, or PNA) polymers could in principle have been the early genetic molecules (Table 7). The self-assembly of XNA polymers from monomers requires polymerization under prebiotic conditions. Most biopolymerization reactions are kinetically unfavorable processes in the absence of specific catalysts such as enzymes, especially in aqueous environments. However, non-aqueous environments, such as a hydrocarbon environment, might shift the polymerization thermodynamics and kinetics sufficiently. Thus a lipid container might be able to mediate XNA polymerization if the XNA is sufficiently lipophilic. This remains to be demonstrated.

A key unresolved question in the origin of life has to do with the early catalytic ability of inheritable molecules. What types of reactions could these early genes have catalyzed? It is more likely that they carried some direct chemical catalysis. In the weakest sense in the context of a protocell, genes are inheritable molecules that control some of the naturally occurring self-assembly and metabolic processes. This concept is at the center of the Los Alamos protocell approach (Rasmussen *et al.*, 2003). In this model, only certain sequences of PNA are assumed to be able to catalyze the reaction of lipid precursors and turn them into surfactant molecules, thus controlling the overall growth of the protocell. For example, templating polymers could serve as the genetic material in the protocell as long as container formation can be controlled by these templating polymers. Thus, chemistry becomes biology when the inheritable molecules control one or more of the container or metabolic functionalities. This means that the templating molecules are catalysts (conceptually similar to ribozymes) that enhance their own fitness within their microenvironment (container + available building blocks) and, thus, their own proliferation.

However, the prebiotic synthesis of XNAs that successively function as the first genetic molecules is questionable because of the structural complexity of these molecules. PAHs might be incorporated more directly into informational polymers, as originally proposed by Platts (*e.g.*, Chapter 17, Hazen, 2005), but we propose a simplified variation on this idea as depicted in Fig. 7, middle. However, an even simpler PAH gene concept can be extracted from what we know about certain polymers: particular types of PAHs interact with each other specifically to form highly ordered liquid crystal phase structures. Hydrogen bonding, π-stacking, and ionic interactions may mediate the formation of these complex structures (Kato, 2002). A template block polymer with a particular sequence of PAHs might be synthesized (Fig. 7, left). Different neighboring complementary PAH blocks would thus allow a block-specific template replication in functional analogy with the well-known Watson–Crick base pairing (Fig. 7, right). Compared with XNA templating, which is based on precise molecular hybridization, the direct PAH block polymer templating proposed in Fig. 7, left, is low resolution, yet potentially highly robust, and may be more easily synthesized in prebiotic environments. Experimental demonstration of such templating processes would be extremely interesting.

Integration of the metabolic, genetic, and containment components

Despite the extensive identified set of possible PAH-based functionalities, it is not at all clear how these components could have self-organized into a minimal living system (see Table 7 and previous subsections). Incompatible component and reaction conditions as well as possible cross-reactions prevent us from simply joining three random structures from the metabolic, genetic, and con-

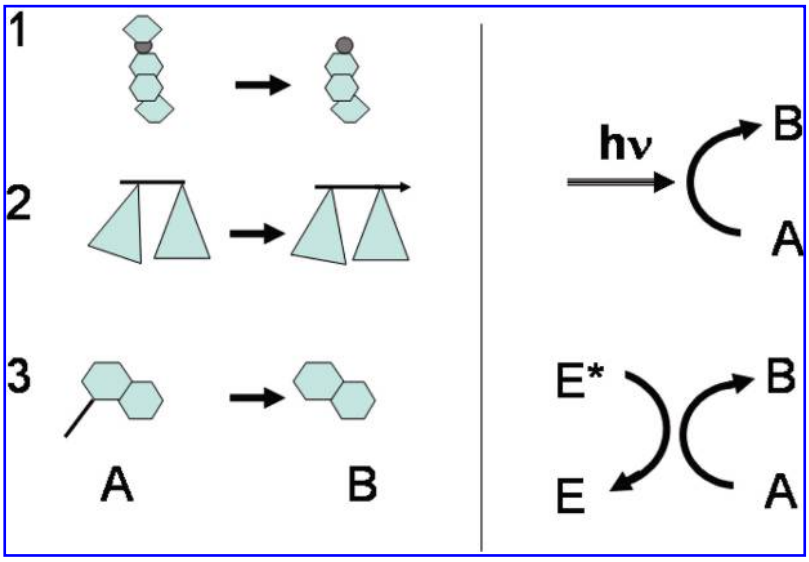


FIG. 6. Metabolic reactions 1, 2, and 3 for the generation of functional container-, gene-, and metabolic-building blocks. Resource components 1A, 2A, and 3A have to be converted into functional building blocks 1B, 2B, or 3B, respectively, through either the formation or the breakage of a covalent bond. The energy can either be provided as redox or as photo energy. One or more of these metabolic reactions have to be mediated within the protocell container. Minimal life requires an autonomous minimal metabolic system; without available free energy to drive the replication process, however simple it might be, no life is possible. Several cooperative feedbacks are possible; *e.g.*, the hydrophobic microenvironment within the container aggregate could mediate one or more of the metabolic processes. Alternatively, a genetic polymer, such as a ribozyme, could catalyze one or more of the metabolic processes.

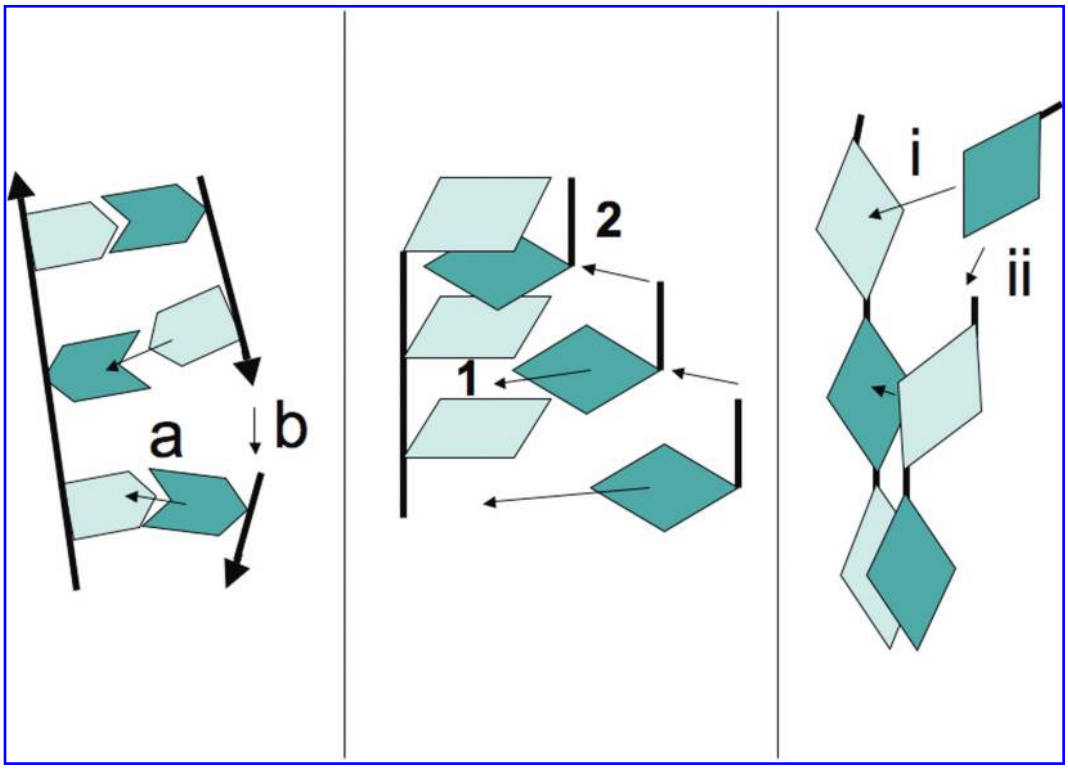


FIG. 7. Some examples of possible early templating polymers. (**Left panel**) The well-known base recognition process where complementary nucleic bases (XNA, *e.g.*, RNA or PNA) align along an existing template (a) and thus mediate the polymerization (b) of a complementary template. (**Middle panel**) Example of π-stacking of particular PAH components [PAH stacking as the basis for an informational polymer was originally proposed by Platts (see Hazen, 2005)] (1), which, in term, could also polymerize (2) to form a complementary template. (**Right panel**) An even simpler PAH π-stacking-based templating (i) and polymerization (ii) could occur based on PAH elements that match complementary PAHs already covalently connected into a string. Compared with the modern XNA hybridization, one should expect this proposed linear PAH templating to be less specific, but more readily synthesized in a prebiotic environment.

tainer sets discussed in the three previous sections. Sufficient conditions for the self-assembly of a (minimal) protocell based on our operational definition are discussed in Fig. 8. A PAH protocell might be able to digest resources and turn them into building blocks that allow the protocell to grow and divide. Furthermore, an imperfect genetic replicator might be able to catalyze part of the digestive, growth, and division processes. A similar organizationally closed protocell structure based on different components is the basis for the Los Alamos protocell (Rasmussen *et al*., 2003). In fact, all of the proposed protocellular component models are faced with this same integration problem (Morowitz *et al*., 1988; Luisi *et al*., 1994; Lee *et al*., 1996; Szostak *et al*., 2001; Pohorille and Deamer, 2002; von Kiedrowski *et al*., 2003). The more theoretically based protocellular approaches traditionally focus on these integration issues, as the theoretical and computational frameworks readily provide a freedom to explore these issues divorced from complex experimental compatibility issues. Many groups have addressed different aspects of these theoretical integration issues both for chemical systems and in more abstract settings mostly based on John von Neumann's (1966) and Manfred Eigen's seminal work (Eigen, 1971) (see, *e.g*., Langton, 1984; Farmer *et al*., 1986; Kauffman, 1986; Rasmussen, 1988; Bollobas and Rasmussen, 1989; Boerlijst and Hogeweg, 1991; Ganti, 1997; Wächtershäuser, 1997; Sayama, 1998; Segré *et al*., 2000). In addition to these rational protocell designs, complementary evolutionary design principles are pursued using computer-controlled microfluidics for life support and complementation as well as for combinatorial screening, which is the basis for the PACE project (see http://www.protocells.org/PACE).

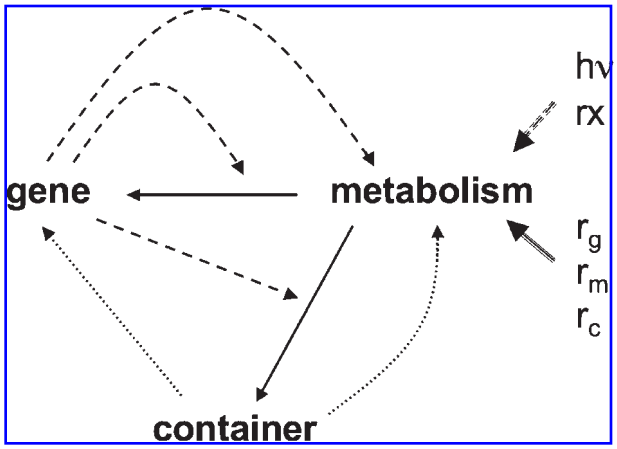


FIG. 8. Sufficient functional interactions that integrate a minimal PAH-based living system (protocell). A lipophilic inheritable replicator (see Possible template chemistry) catalyzes (dashed arrows) the metabolic formation (solid arrows) of both gene- and container-building blocks (see Possible metabolic chemistry). The amphiphilic container molecules self-assemble and increase the container size (see Possible container chemistry). The container ensures a high local concentration and facilitates thermodynamic reaction conditions (dotted arrows) of both the hydrophobic metabolic molecules and the amphiphilic replicator polymers. The necessary free energy can either be redox- or photo-based. Ideally, all three protocellular components should be internally synthesized or directly available from the environment: the resources for the gene monomers, r_g, the metabolic complexes, r_m, and the container monomers, r_c.

Besides the obvious necessary chemical compatibility for component integration, a functional compatibility is necessary as well. Without some coordination in the component formation kinetics, destructive component imbalances may occur. In the Chemoton model of Ganti (2004), the proposed solution is detailed stoichiometric control of the components. Molecular control is imposed such that for each molecule of type *X* synthesized, a fixed number, *y*, of molecules of type *Y* are synthesized, while another fixed number, *z*, of molecules of type *Z* are synthesized, etc. (Munteanu and Sole, 2005). In contrast, the Los Alamos protocell (Rasmussen *et al*., 2003) defines autocatalytic control between the components, where relative concentrations regulate the component balance. Although catalytic feedback is less restrictive than stoichiometric control, both methods should work, in principle (Rocheleau *et al*., 2006). However, most published protocell models are not explicit as to how the integrative feedback structures are ensured and, rather, focus more on the experimental issues related to the key protocellular reactions such as template-directed replication and container replication. The insurance of mutual cooperativity between the three main components is the central and likely most difficult experimental protocellular research problem. So far this has not yet been obtained experimentally with any protocell design.

Little is known about the details of the transition from nonliving to living matter on Earth discussed in Fig. 9 or the historical events surrounding it. It is not known whether only one transition occurred, or several transitions converged into a single chemical "solution" (pathway 1a and b in Fig. 9). It is possible that many transitions (pathways 1, 2, and 3, Fig. 9) could be constructed in the laboratory, some of which

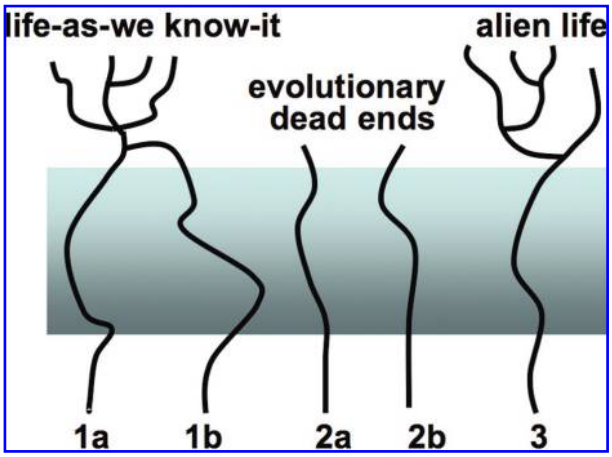


FIG. 9. Many transitions may exist between nonliving and living matter based on our definition of minimal life. To date we know that at least one such transition must have occurred that eventually led to modern life. We propose a PAH-based transition as the most likely candidate. Whether it was a single transition or several transitions that converged into life-as-we-know-it is unknown (pathways 1a and 1b). We hypothesize that many transitions could be constructed in the laboratory that may be compatible with the natural origins of life. Though most of the recently published protocell designs could satisfy the definition of minimal life, they may not be able to evolve any further and thus be unable to converge toward contemporary life. Such evolutionary blind alleys (pathways 2a and 2b) would still provide us with critical information about what life is, its origin, and possible occurrence elsewhere. Finally, it might be possible to create life forms based on chemistry and an organization that is significantly different from life-as-we-know-it, such that truly alien life could evolve from it (pathway 3). In any event, we propose that studying this transition should be a main focus area of astrobiology.

were alive but could not evolve (pathway 2a and b, Fig. 9). These would then be evolutionary blind alleys, living systems that could only evolve slightly modified simple functionalities but not significant novel functions. Many such evolutionarily limited self-replicating systems have been explored in simulation (Ray, 1991; Lindgren, 1992; Lindgren and Nordahl, 1994; Adami, 1995; Linski *et al*., 1999). What the minimal system requirements are for open-ended evolution is a key open research question for evolutionary dynamics (Bedau *et al*., 2000). Today, we have only two documented open-ended evolutionary processes: biological evolution and human technological evolution (Bedau *et al*., 1998; Skusa and Bedau, 2002).

Different transitions from non-living to living matter should be possible, and alternative routes should be pursued experimentally. It may even be possible to construct living systems based on alternative chemical pathways and organizational principles, which are evolvable. This would be of immense scientific interest, but presently no obvious alternatives exist to carbon-based (bio)chemistry and a cooperation among genetics, metabolism, and container (see, *e.g*., Bains, 2004; Davies and Lineweaver, 2005). We therefore propose to focus on carbon chemistry. We further propose that the key problem of the origins of life should be tackled experimentally in a step-by-step manner by (1) initially identifying functional container, metabolic, and genetic components followed by (2) a cooperative integration of the three components. At the present time, the astrobiology community has not sufficiently addressed experimentally the critical transition from non-living to living matter.

It may be possible to address several key problems in the origins of life in a rational manner by investigating *PAH-based protocell components*, which are abundantly distributed throughout our galaxy (see Table 7) and may have the necessary functionalities (stability and flexibility) to serve as the building blocks of life. Why not experimentally explore the most abundant organic molecules on the young Earth as the building blocks for life, aromatic carbonaceous matter?

Proposed experiments

Many direct tests for container formation are readily possible by studying solvation and self-assembly of a variety of starting materials. For example, the formation of vesicles from different starting materials, including carboxylic acids (Apel *et al*., 2002) and/or PAH derivatives, should be investigated, and their stability under simulated early Earth conditions such as ocean or subsurface pore space habitats studied. This would provide important constraints for understanding compartmentalization on the young Earth.

Also, simple metabolic processes can directly be tested, *e.g*., based on PAH-mediated surfactant production. Here we discuss two examples that couple simple prebiotic metabolic processes with the production of container materials. Consider oxidation of hydrocarbons to amphiphilic molecules driven by PAH derivatives. PAHs have larger electronic delocalization structures (increasing aromaticity) and are stable at the ground state in the dark. However, they can absorb light at

longer wavelengths (>350 nm, visible light) because of the delocalization structures and, thus, can be easily raised to excited states that have higher oxidation potentials. Consequently, PAHs can photo-oxidize some hydrocarbons to amphiphiles. For example, by using 2-ethyl anthracene, dodecane can be converted into a mixture of dodecanone and dodecanol, both of which are surface-active amphiphiles. Thus, amphiphiles can be generated from hydrocarbons using PAHs and light as an energy source (Klein and Pilpel, 1973; Deamer, 1992).

Another possible metabolic process generates amphiphilic molecules by PAH-sensitized photo-oxidation. It has been demonstrated that light energy can use low concentrations of oxygen to convert long-chain olefins or PAHs to their polar photo-oxidation products through PAH sensitization reactions. We have shown that abietic diterpene derivatives can be efficiently oxidized to hydroxy or carbonyl abietic derivatives by perylene photosensitization (Chen *et al*., 1998b). Significantly, the reaction does not proceed through the conventional photo-oxidation mechanism, which involves singlet oxygen formation, but rather by a novel oxygen interception of a photogenerated exciplex to yield the superoxide. Since many fused PAHs have properties similar to perylene, PAH-sensitized oxidation of long-chain alkenes is a plausible pathway for the production of amphiphilic compounds using trace molecular oxygen available on the prebiotic young Earth.

The possible formation of simple PAH-based informational (templating) polymers can also readily be tested. Certain types of PAHs interact specifically with each other to form highly ordered liquid crystal phase structures. For example, molecules such as triphenylene, porphyrin, phthalocyanine, coronene, and other aromatic molecules can form discotic liquid crystals, in which molecules are stacked one-dimensionally and assemble into cylindrical structures because of the π–π electronic interactions. However, the formation of liquid crystalline phases can also be achieved by hydrogen bonding between amphiphilic molecules bearing complementary functional groups. For example, diaminopyridine and uracil derivatives with long aliphatic tails form liquid crystals (Bricnne *et al*., 1989). The strong tendency for molecules to form liquid crystal phases renders a cluster of such molecules that can specifically interact with another cluster of similar molecules. Such block–block interactions that form a larger liquid crystal phase will provide an avenue for PAH-based informational templating. Accordingly, a template block polymer with a particular sequence of PAHs with π stacking block and hydrogen binding stacking block, respectively, would allow block specific recognitions and hybridization.

To the best of our knowledge, the example described herein is the first suggestion of the preparation of PAH-based templating and hybridization through molecular recognition of block motifs. Further study of the structures of interacting components that lead to stable systems is required. In addition to proper geometry and complementarity, the interacting blocks must also benefit cooperatively, which includes secondary molecular interactions, *e.g*., similar to the secondary molecular interactions within a DNA duplex that forms and supports the helix structure.

Finally, the structural and chemical changes of aromatic macromolecular matter (as found in carbonaceous meteorites) when exposed to thermal and radiation processing or oxidation need to be investigated in the laboratory. It is important to understand how such material fragments and whether subunits can act as functional building blocks in the origin of life.

The materials used should be similar to compounds found in meteorites, or similar to the possible inventory of early organic compounds. Obviously, a variety of practical issues emerge with this proposal. PAHs are poorly soluble in water, a characteristic that may cause experimental obstacles. However, a variety of organic solvents that could simulate prebiotic hydrocarbons can be used. Also, the hydrophobic nature of PAHs enables them to concentrate in micelles and other aggregates, which could be an advantage, *e.g*., if the aggregates can be used as reaction containers. The molecular diversity in the extraterrestrially delivered PAHs could also cause problems. However, solutions to these problems may be discovered, as it is already known that many PAHs can function as container elements and energy transduction elements, as well as templating genetic components. Our hypothesis is that it may be possible to assemble a functional protocell using aromatic material.

CONCLUSION

Carbon-based life as we know it has adapted to a wide range of extreme environments on

planet Earth. Prebiotic soups resulting from atmospheric synthesis, hydrothermal vents, synthesis from outgassing, and extraterrestrial infall of organic material have been extensively discussed in the last decades (Miller, 1953; Chyba and Sagan, 1992; Russell and Hall, 1997; Gold, 1999; Ehrenfreund *et al*., 2002). The crucial step in life's origin, the successful assembly of functional components, must be strongly determined by the environment and the ability of organic materials to be concentrated within it (Cleaves and Chalmers, 2004). Besides the presence of organic precursor compounds, water, energy, and catalytic surfaces are among the vital conditions necessary to form more complex structures. While the prebiotic synthesis of many of the constituents of modern biochemical systems has been elucidated, their assembly into a living system remains unsolved. It is possible that chemists will succeed in producing synthetic life using chemicals alien to biochemistry, and it is possible that the earliest terrestrial life was composed of compounds no longer used in modern biochemistry. Precursor molecules that make up our genetic material, such as amino acids, nucleobases, and sugars, have attracted a lot of attention in prebiotic chemistry. However, their stability with respect to the harsh temperature and radiation conditions on the young Earth was very limited. Therefore, it remains an open question as to whether those compounds have been used from the very start to build up life.

In this paper, we have combined knowledge from astrochemistry, prebiotic chemistry, and artificial life in an attempt to construct a model whereby life could have evolved from a different starting material, namely, aromatic hydrocarbons. PAHs and aromatic macromolecules are the dominant organic material in space. Such material may also have been produced terrestrially, and more efficiently than the typical compounds found in modern biochemistry. Aromatic material is extremely stable and quite versatile, and could fulfill different functions by side-group addition and polymerization.

We have elaborated how PAHs might function as container elements, energy transduction elements, and templating genetic components. Our hypothesis is that it is possible to assemble a functional protocell using predominantly aromatic material. We have outlined in Functional Assemblies and Their Integration into a Minimal Life Form how a variety of transitions from nonliving to living matter could be experimentally tested.

Organic chemistry in space seems to follow common pathways throughout the universe. Carbonaceous molecules in the gas or solid state (such as dusty or icy grains) are observed in similar abundances and composition in numerous galaxies, including our own, as discussed in The Formation and Evolution of Organic Material in Space. Aromatic material in the gas phase and in macromolecular form makes up most of the carbon in the interstellar medium, in comets, and in meteorites as well. It is likely that aromatic hydrocarbons are the most abundant organic materials delivered to early planets (Ehrenfreund *et al*., 2002). Although we cannot exclude the possibility of life based on alien chemistry using neither carbon nor water, the ubiquity of carbon, water, and aromatics across the universe makes it tempting to focus on an origin of life based on aqueous carbon chemistry.

It seems plausible to us that the varied chemistry of diverse macromolecular and aromatic carbon could have played an important role in the origin of life, especially given the stability and likely abundance of such materials on primitive planets. It is even possible that this material was used as the key elements to assemble life, which is why we now propose the "Aromatic World."

ACKNOWLEDGMENTS

P.E. acknowledges NWO grant 016.023.010. S.R. and L.C. acknowledge Los Alamos LDRD-DR grant on Protocell Assembly and the European Commission grant on Programmable Artificial Cell Evolution (PACE). We thank Nick Cox for graphic support and Mark Sephton for discussion.

ABBREVIATIONS

FTT, Fischer–Tropsch type; GNA, glycol nucleic acid; HMP, hyper-metal-poor; PAH, polycyclic aromatic hydrocarbon; PNA, peptide nucleic acid; TNA, threose nucleic acid; UV, ultraviolet.

REFERENCES

Adami, C. (1995) Learning and complexity in genetic auto-adaptive systems. *Physica D* 80, 154–170.

Allamandola, L.J., Hudgins, D.M., and Sandford, S.A. (1999) Modeling the unidentified infrared emission

with combinations of polycyclic aromatic hydrocarbons. *Astrophys. J.* 511, 115–119.

Apel, C.L., Deamer, D.W., and Mautner, M.N. (2002) Self-assembled vesicles of monocarboxylic acids and alcohols: conditions for stability and for the encapsulation of biopolymers. *Biochim. Biophys. Acta* 1559, 1–9.

Bada, J., Bigham, C., and Miller, S. (1994) Impact melting of frozen oceans on the early Earth: implications for the origin of life. *Proc. Natl. Acad. Sci. USA* 91, 1248–1250.

Bains, W. (2004) Many chemistries could be used to build living systems. *Astrobiology* 4, 137–167.

Basiuk, V. and Navarro-Gonzalez, R. (1996) Dust particles in the atmospheres of terrestrial planets and their roles for prebiotic chemistry: an overview. *Astrophys. Space Sci.* 236, 61–75.

Beaulieu, J.P., Bennett, D.P., Fouque, P., Williams, A., Dominik, M., Jorgensen, U.G., Kubas, D., Cassan, A., Coutures, C., Greenhill, J., Hill, K., Menzies, J., Sackett, P.D., Albrow, M., Brillant, S., Caldwell, J.A., Calitz, J.J., Cook, K.H., Corrales, E., Desort, M., Dieters, S., Dominis, D., Donatowicz, J., Hoffman, M., Kane, S., Marquette, J.B., Martin, R., Meintjes, P., Pollard, K., Sahu, K., Vinter, C., Wambsganss, J., Woller, K., Horne, K., Steele, I., Bramich, D.M., Burgdorf, M., Snodgrass, C., Bode, M., Udalski, A., Szymanski, M.K., Kubiak, M., Wieckowski, T., Pietrzynski, G., Soszynski, I., Szewczyk, O., Wyrzykowski, L., Paczynski, B., Abe, F., Bond, I.A., Britton, T.R., Gilmore, A.C., Hearnshaw, J.B., Itow, Y., Kamiya, K., Kilmartin, P.M., Korpela, A.V., Masuda, K., Matsubara, Y., Motomura, M., Muraki, Y., Nakamura, S., Okada, C., Ohnishi, K., Rattenbury, N.J., Sako, T., Sato, S., Sasaki, M., Sekiguchi, T., Sullivan, D.J., Tristram, P.J., Yock, P.C., and Yoshioka, T. (2006) Discovery of a cool planet of 5.5 Earth masses through gravitational microlensing. *Nature* 439, 437–440.

Beck, A., Lohrmann, R., and Orgel, L.E. (1967) Phosphorylation with inorganic phosphates at moderate temperatures. *Science* 157, 952–952.

Becker, L. and Bunch, T.E. (1997) Fullerenes, fulleranes and PAHs in the Allende meteorite. *Meteoritics* 32, 479–487.

Bedau, M.A., Snyder, E., and Packard, N.H. (1998) A classification of long-term evolutionary dynamics. In *Artificial Life VI*, edited by C. Adami, R. Belew, H. Kitano, and C. Taylor, MIT Press, Cambridge, MA, pp. 228–237.

Bedau, M.A., McCaskill, J.S., Packard, N.H., Rasmussen, S., Adami, C., Green, D.G., Ikegami, T., Kaneko, K., and Ray, T. (2000) Open problems in artificial life. *Artif. Life* 6, 363–376.

Beers, T.C. (2005) Where are we now, and where are we going? In *ASP Conference Series, Vol. 336: Cosmic Abundances as Records of Stellar Evolution and Nucleosynthesis in Honor of David L. Lambert*, edited by T.G. Barnes III and F.N. Bash, Astronomical Society of the Pacific, San Francisco, pp. 283–289.

Beier, M., Reck, F., Wagner, T., Krishnamurthy, R., and Eschenmoser, A. (1999) Chemical etiology of nucleic acid structure: comparing pentopyranosyl-(2′ → 4′) oligonucleotides with RNA. *Science* 283, 699–703.

Berndt, M.E., Allen, D.E., and Seyfried, W.E. (1996) Reduction of CO_2 during serpentinization of olivine at 300°C and 500 bar. *Geology* 24, 351–354.

Bernstein, M.P., Dworkin, J.P., Sandford, S.A., Cooper, G.W., and Allamandola, L.J. (2002) Racemic amino acids from the ultraviolet photolysis of interstellar ice analogues. *Nature* 416, 401–403.

Bertoldi, F., Cox, P., Neri, R., Carilli, C.L., Walter, F., Omont, A., Beelen, A., Henkel, C., Fan, X., Strauss, M.A., and Menten, K.M. (2003) High-excitation CO in a quasar host galaxy at z = 6.42. *Astron. Astrophys.* 409, L47–L50.

Blum, J. (2004) Grain growth and coagulation. In *ASP Conference Series 309: The Astrophysics of Dust*, edited by A.N. Witt, G.C. Clayton, and B.T. Draine, Astronomical Society of the Pacific, San Francisco, pp. 369–392.

Bockelée-Morvan, D., Crovisier, J., Mumma, M.J., and Weaver, H.A. (2004) The composition of cometary volatiles. In *COMETS II*, edited by M. Festou, H.U. Keller, and H. Weaver, University of Arizona Press, Tucson, pp. 391–423.

Boerlijst, M. and Hogeweg, P. (1991) Self-structuring and selection: spiral waves as a substrate for prebiotic evolution. In *Santa Fe Institute Studies in the Sciences of Complexity, Vol. X: Artificial Life II*, edited by C. Langton, C. Taylor, J.D. Farmer, and S. Rasmussen, Addison-Wesley, Redwood, CA, pp. 255–276.

Bollobas, B. and Rasmussen, S. (1989) First cycles in random directed graph processes. *Discrete Math.* 75, 55–68.

Boss, A.P. (2004) From molecular clouds to circumstellar disks. In *COMETS II*, edited by M. Festou, H.U. Keller, and H. Weaver, University of Arizona Press, Tucson, pp. 67–80.

Botta, O. and Bada, J. (2002) Extraterrestrial organic compounds in meteorites. *Geophys. Survey* 23, 411–467.

Bricnne, M.J., Galard, I., Lehn, J.M., and Stihor, J. (1989) Macroscopic expression of molecular recognition. Supramolecular liquid crystalline phase induced by association of complementary heterocyclic components. *J. Chem. Soc. Chem. Commun.* 14, 1868–1877.

Brooks, R. (2001) The relationship between matter and life. *Nature* 409, 409–411.

Butlerov, A. (1861) Formation synthétique d'une substance sucreé. *C.R. Acad. Sci.* 53, 145–147.

Cairns-Smith, G. (2005) Sketches for a mineral genetic material. *Elements* 1, 157–161.

Canuto, V., Levine, J., Augustsson, T., and Imhoff, C. (1983) Oxygen and ozone in the early Earth's atmosphere. *Precambrian Res.* 20, 109–120.

Cataldo, F. (2004) From elemental carbon to complex macromolecular networks in space. In *Astrophysics and Space Science Library, Vol. 305: Astrobiology: Future Perspectives*, edited by P. Ehrenfreund, W.M. Irvine, T. Owen, L. Becker, J. Blank, J.R. Brucato, L. Colangeli, S. Derenne, A. Dutrey, D. Despois, A. Lazcano, and F. Robert, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 97–126.

Cernicharo, J., Heras, A. Tielens, A.G.G.M., Pardo, J.R., Herpin, F., Guèlin, M., and Waters, L.B.F.M. (2001) In-

frared Space Observatory's discovery of C_4H_2, C_6H_2, and benzene in CRL 618. *Astrophys. J.* 546, 123–126.
Chadha, M. and Choughuley, A. (1984) Synthesis of prebiotic molecules—role of some carbonyl compounds in prebiotic chemistry. *Orig. Life* 14, 469–476.
Charnley, S., Ehrenfreund, P., and Kuan, Y. (2003) Molecules in space. *Physics World Oct*, 35–38.
Chen, L., Lucian, L., and Whitten, D.G. (1998a) Cooperative electron transfer fragmentation reactions. Amplification of a photoreaction through a tandem chain fragmentation of acceptor and donor pinacols. *J. Am. Chem. Soc*. 120, 439–440.
Chen, L., Lucia, L.A., Gaillard, E.R., Whitten, D.G., Icil, H., and Icli, S. (1998b) Photooxidation of a conjugated diene by an exciplex mechanism: amplification via radical chain reactions in the perylene diimide-photosensitized oxidation of α-terpinene. *J. Phys. Chem. A* 102, 9095–9098.
Chen, L., Geiger, C., Perlstein, J., and Whitten, D.G. (1999) Self-assembly of styryl naphthalene amphiphiles in aqueous dispersions and interfacial films: aggregate structure, assembly properties, photochemistry and photophysics. *J. Phys. Chem. B* 103, 9161–9167.
Cherchneff, I., Barker, J., and Tielens, A. (1992) Polycyclic aromatic hydrocarbon formation in carbon-rich stellar envelopes. *Astrophys. J.* 401, 269–287.
Chick, K. and Cassen, P. (1997) Thermal processing of interstellar dust grains in the primitive solar environment. *Astrophys. J*. 477, 398–409.
Choughuley, A., Subbaraman, A., Kazi, Z., and Chadha, M. (1977) A possible prebiotic synthesis of thymine: uracil-formaldehyde-formic acid reaction. *BioSystems* 9, 73–80.
Chyba, C. and Sagan, C. (1992) Endogenous production, exogenous delivery and impact-shock synthesis of organic molecules: an inventory for the origins of life. *Nature* 355, 125–132.
Chyba, C., Thomas, P., Brookshaw, L., and Sagan, C. (1990) Cometary delivery of organic molecules to the early Earth. *Science* 249, 366–373.
Cleaves, H. and Chalmers, J. (2004) Extremophiles may be irrelevant to the origin of life. *Astrobiology* 4, 1–9.
Clemett, S., Maechling, C., Zare, R., Swan, P., and Walker, R. (1993) Identification of complex aromatic molecules in individual interplanetary dust particles. *Science* 262, 721–725.
Cody, G.D. and Alexander, C.M.O.' (2005) NMR studies of chemical structural variation of insoluble organic matter from different carbonaceous chondrite groups. *Geochim. Cosmochim. Acta* 69, 1085–1097.
Cody, G., Boctor, N.Z., Filley, T.R., Hazen, R.M., Scott, J.H., Sharma, A., and Yoder, H.S. (2000) Primordial carbonylated iron-sulfur compounds and the synthesis of pyruvate. *Science* 289, 1337–1340.
Colgate, S., Rasmussen, S., Solem, J., and Lackner, K. (2003) An astrophysical theory for a universal origin of life. *Adv. Complex Syst*. 6, 487–505.
Cooper, G., Onwo, W.M., and Cronin, J.R. (1992) Alkyl phosphonic acids and sulfonic acids in the Murchison meteorite. *Geochim. Cosmochim. Acta* 56, 4109–4115.
Corliss, J., Dymond, J., Gordon, L., Edmond, J., von Herzen, R., Ballard, R., Green, K., Williams, D., Bainbridge, A., Crane, K., and van Andel, T. (1979) Submarine thermal springs on the Galapagos Rift. *Science* 203, 1073–1083.
Corliss, J., Baross, J., and Hoffman, S. (1981) An hypothesis concerning the relationship between submarine hot springs and the origin of life on Earth. *Oceanol. Acta* 4 (Suppl.), 59–69.
Crovisier, J. (2004) The molecular complexity of comets. In *Astrophysics and Space Science Library, Vol. 305: Astrobiology: Future Perspectives*, edited by P. Ehrenfreund, W.M. Irvine, T. Owen, L. Becker, J. Blank, J.R. Brucato, L. Colangeli, S. Derenne, A. Dutrey, D. Despois, A. Lazcano, and F. Robert, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp.179–204.
Daigne, F., Olive, K.A., Vangioni-Flam, E., Silk, J., and Audouze, J. (2004) Cosmic star formation, reionization, and constraints on global chemical evolution. *Astrophys. J*. 617, 693–706.
Dartois, E., Muñoz Caro, G.M., Deboffle, D., Montagnac, G., and D'Hendecourt, L. (2005) Ultraviolet photoproduction of ISM dust. Laboratory characterisation and astrophysical relevance. *Astron. Astrophys*. 432, 895–908.
Davies, P.C.W. and Lineweaver, C.H. (2005) Finding a second sample of life on earth. *Astrobiology* 5, 154–163.
Deamer, D. (1985) Boundary structures are formed by organic components of the Murchison carbonaceous chondrite. *Nature* 317, 792–794.
Deamer, D.W. (1992) Polycyclic aromatic hydrocarbons: primitive pigment systems in the prebiotic environment. *Adv. Space. Res*. 12, 183–189.
De Graaf, R.M. and Schwartz, A.W. (2005) Thermal synthesis of nucleoside H-phosphonates under mild conditions. *Orig. Life Evol. Biosph*. 35, 1–10.
Discher, D.E. and Eisenberg, A. (2002) Polymer vesicles. *Science* 297, 967–973.
Duley, W.W. and Lazarev, S. (2004) Ultraviolet absorption in amorphous carbons: polycyclic aromatic hydrocarbons and the 2175 Å extinction feature. *Astrophys. J*. 612, 33–35.
Ehrenfreund, P. and Charnley, S.B. (2000) Organic molecules in the ISM, comets and meteorite: a voyage from dark clouds to the early Earth. *Annu. Rev. Astron. Astrophys*. 38, 427–483.
Ehrenfreund, P. and Fraser, H. (2003) Ice chemistry in space. In *NATO ASI Series: Solid State Astrochemistry*, edited by V. Pirronello and J. Krelowski, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 317–356.
Ehrenfreund, P. and Martin, E.L. (2006) Astronomical constraints for life on terrestrial planets. In *COLE Series, Vol. 10: Life as We Know It as: Cellular Origins, Life in Extreme Habitats and Astrobiology*, Springer Verlag, New York, in press.
Ehrenfreund, P., Robert, F., d'Hendecourt, L., and Behar, F. (1991) Comparison of interstellar and meteoritic organic matter at 3.4 μm. *Astron. Astrophys*. 252, 712–717.
Ehrenfreund, P., Irvine, W., Becker, L., Blank, J., Brucato, J., Colangeli, L., Derenne, S., Despois, D., Dutrey, A.,

Fraaije, H., Lazcano, A., Owen, T., and Robert, F. (2002) Astrophysical and astrochemical insights into the origin of life. *Rep. Prog. Phys*. 65, 1427–1487.

Ehrenfreund, P., Charnley, S., and Wooden, D. (2004) From ISM material to comet particles and molecules. In *COMETS II*, edited by M. Festou, H.U. Keller, and H. Weaver, University of Arizona Press, Tucson, pp. 115–133.

Eigen, M. (1971) Self-organization of matter and the evolution of macromolecules. *Naturwissenshaften* 58, 465–523.

Eriksson, M., Christensen, L., Schmidt, J., Haaima, G., Orgel, L., and Nielsen, P. (1998) Sequence dependent N-terminal rearrangement and degradation of peptide nucleic acid (PNA) in aqueous solution. *New J. Chem*. 22, 1055–1059.

Eschenmoser, A. (2004) The TNA-family of nucleic acid systems: properties and prospects. *Orig. Life Evol. Biosph*. 34, 277–306.

Farmer, J.D., Kauffman, S., and Packard, N.H. (1986) Autocatalytic replication of polymers. *Physica D* 22, 50–67.

Ferris, J. (1992) Chemical markers of prebiotic chemistry in hydrothermal systems. *Orig. Life Evol. Biosph*. 22, 109–134.

Flynn, G.J., Keller, L.P., Feser, M., Wirick, S., and Jacobsen, C. (2003) The origin of organic matter in the solar system: evidence from the interplanetary dust particles. *Geochim. Cosmochim. Acta* 67, 4791–4806.

Foing, B.H. and Ehrenfreund, P. (1994) Detection of two interstellar absorption bands coincident with spectral features of C_{60}^+. *Nature* 369, 296–298.

Foing, B.H. and Ehrenfreund, P. (1997) New evidences for interstellar C_{60}^+. *Astron. Astrophys*. 317, 59–62.

Frenklach, M. and Feigelson, E.D. (1989) Formation of polycyclic aromatic hydrocarbons in circumstellar envelopes. *Astrophys. J*. 341, 372–384.

Fuller, W., Sanchez, R., and Orgel, L. (1972) Prebiotic synthesis. VII. Solid-state synthesis of purine nucleosides. *J. Mol. Evol*. 1, 249–257.

Ganti, T. (1997) Biogenesis itself. *J. Theor. Biol*. 187, 583–593.

Ganti, T. (2004) *Chemoton Theory*, Kluwer Academic Publisher, Dordrecht, The Netherlands.

Gardinier, A., Derenne, S., Robert, F., Behar, F., Largeau, C., and Maquet, J. (2000) Solid state CP/MAS ^{13}C NMR of the insoluble organic matter of the Orgueil and Murchison meteorites: quantitative study. *Earth Planet. Sci. Lett*. 184, 9–21.

Gesteland, R., Cech, T., and Atkins, J. (1999) *The RNA World*, 2nd ed., Cold Spring Harbor Laboratory Press, Cold Spring Harbor, New York.

Goeres, A. and Sedlmayr, E. (1992) The envelopes of R Coronae Borealis stars. I—A physical model of the decline events due to dust formation. *Astron. Astrophys*. 265, 216–236.

Gold, T. (1999) *The Deep Hot Biosphere*, Copernicus, New York.

Greenberg, J. (1998) Making a comet nucleus. *Astron. Astrophys*. 330, 375–380.

Guillois, O., Ledoux, G., and Reynaud, C. (1999) Diamond infrared emission bands in circumstellar media. *Astrophys. J*. 521, 133–136.

Hayatsu, R., Studier, M., and Anders, E. (1971) Origin of organic matter in the early solar system. IV. Amino acids. Confirmation of catalytic synthesis by mass spectrometry. *Geochim. Cosmochim. Acta* 35, 939–951.

Hayatsu, R., Studier, M., Matsuoka, S., and Anders, E. (1972) Origin of organic matter in early solar system. VI. Catalytic synthesis of nitriles, nitrogen bases, and porphyrin-like pigments. *Geochim. Cosmochim. Acta* 36, 555–571.

Hazen, R.M. (2005) *Genesis: The Scientific Quest for Life's Origins*, Joseph Henry Press, Washington, DC.

Henning, T. and Salama, F. (1998) Carbon in the Universe. *Science* 282, 2204–2210.

Henning, Th., Jäger, C., and Mutschke, H. (2004) Laboratory studies of carbonaceous dust analogs astrophysics of dust. In *ASP Conference Series 309: The Astrophysics of Dust*, edited by A.N. Witt, G.C. Clayton, and B.T. Draine, Astronomical Society of the Pacific, San Francisco, pp. 603–628.

Hill, H. and Nuth, J. (2003) The catalytic potential of cosmic dust: implications for prebiotic chemistry in the solar nebula and other protoplanetary systems. *Astrobiology* 3, 291–304.

Hodgson, G. and Ponnamperuma, C. (1968) Prebiotic porphyrin genesis. Porphyrins from electric discharge in methane, ammonia, and water vapor. *Proc. Natl. Acad. Sci. USA* 59, 22–28.

Holm, N.G. and Andersson, E.M. (1998) Hydrothermal systems. In *Molecular Origins of Life*, edited by A. Brack, Cambridge University Press, Cambridge, UK, pp. 86–99.

Holm, N. and Charlou, J. (2001) Initial indications of abiotic formation of hydrocarbons in the Rainbow ultramafic hydrothermal system, Mid-Atlantic Ridge. *Earth Planet. Sci. Lett*. 191, 1–8.

Horowitz, N. (1945) The evolution of biochemical syntheses. *Proc. Natl. Acad. Sci. USA* 31, 153–157.

Hotani, H., Lahoz-Beltra, R., Combs, B., Hameroff, S., and Rasmussen, S. (1992) Liposomes, microtubules, and artificial cells. *Nanobiology* 1, 61–74.

Huber, C. and Wächtershäuser, G. (1998) Peptides by activation of amino acids with CO on (Ni, Fe)S surfaces and implications for the origin of life. *Science* 281, 670–672.

Iglesias-Groth, S. (2004) Fullerenes and buckyonions in the interstellar medium. *Astrophys. J*. 608, 37–40.

Ingar, A., Luke, R., Hayter, B., and Sutherland, J. (2003) Synthesis of cytidine ribonucleotides by stepwise assembly of the heterocycle on a sugar phosphate. *ChemBioChem* 4, 504–507.

Irvine, W., Crovisier, J., Fegley, B., and Mumma, M.J. (2000) Comets: a link between interstellar and nebular chemistry. In *Protostars and Planets IV*, edited by V. Mannings, A. Boss, and S. Russell, University of Arizona Press, Tucson, pp. 1159–1200.

Iwamoto, N., Umeda, H., Tominaga, N., Nomoto, K., and Maeda, K. (2005) The first chemical enrichment in the Universe and the formation of hyper metal-poor stars. *Science* 309, 451–453.

Joyce, G.F. (2002) The antiquity of RNA-based evolution. *Nature* 418, 214–221.

Joyce, G.F., Schwartz, A.W., Miller, S.L., and Orgel, L.E. (1987) The case for an ancestral genetic system involving simple analogs of the nucleotides. *Proc. Natl. Acad. Sci. USA* 84, 4398–4402.

Kaneda, H., Onaka, T., and Sakon, I. (2005) Detection of PAH emission features from nearby elliptical galaxies with the Spitzer infrared spectrograph. *Astrophys. J.* 632, 83–86.

Kasting, J. (1993) Earth's early atmosphere. *Science* 259, 920–926.

Kasting, J. and Catling, D. (2003) Evolution of a habitable planet. *Annu. Rev. Astron. Astrophys*. 41, 429–463.

Kato, T. (2002) Self-assembly of phase-segregated liquid crystal structures. *Science* 295, 2414–2418.

Kauffman, S. (1986) Autocatalytic sets of proteins. *J. Theor. Biol*. 119, 1–24.

Keefe, A. and Miller, S. (1995) Are polyphosphates or phosphate esters prebiotic reagents? *J. Mol. Evol*. 41, 693–702.

Keefe, A., Miller, S., McDonald, G., and Bada, J.L. (1995) Investigation of the prebiotic synthesis of amino acids and RNA bases from CO_2 using FeS/H_2S as a reducing agent. *Proc. Natl. Acad. Sci. USA* 92, 11904–11906.

Kelley, D.S., Baross, J.A., and Delaney, J.R. (2002) Volcanoes, fluids, and life at mid-ocean ridge spreading centers. *Annu. Rev. Earth Planet. Sci*. 30, 385–491.

Kelley, D.S., Karson, J.A., Fruh-Green, G.L., Yoerger, D.R., Shank, T.M., Butterfield, D.A., Hayes, J.M., Schrenk, M.O., Olson, E.J., Proskurowski, G., Jakuba, M., Bradley, A., Larson, B., Ludwig, K., Glickson, D., Buckman, K., Bradley, A.S., Brazelton, W.J., Roe, K., Elend, M.J., Delacour, A., Bernasconi, S.M., Lilley, M.D., Baross, J.A., Summons, R.E., and Sylva, S.P. (2005) A serpentinite-hosted ecosystem: the Lost City hydrothermal field. *Science* 307, 1428–1434.

Klein, A.E. and Pilpel, N. (1973) Oxidation of n-alkanes photosensitized by 1-naphthol. *J. Chem. Soc. Faraday I* 69,1729–1736.

Kroto, H.W., Heath, J.R., O'Brien, S.C., Curl, R.F., and Smalley, R.E. (1985) C60: buckminsterfullerene. *Nature* 318, 162–163.

Kuan, Y., Charnley, S.B., Huang, H., Tseng, W., and Kisiel, Z. (2003) Interstellar glycine. *Astrophys. J.* 593, 848–867.

Kuhn, W. and Atreya, S. (1979) Ammonia photolysis and the greenhouse effect in the primordial atmosphere of the earth. *Icarus* 37, 207–213.

Kwok, S. (2004) The synthesis of organic and inorganic compounds in evolved stars. *Nature* 430, 985–991.

Langton, C.G. (1984) Self-reproduction in cellular automata. *Physica D* 10, 135–144.

Larralde, R., Robertson, M., and Miller, S. (1995) Rates of decomposition of ribose and other sugars: implications for chemical evolution. *Proc. Natl. Acad. Sci. USA* 92, 8158–8160.

Lee, D., Granja, J., Martinez, J., Severin, K., and Ghadiri, M. (1996) A self-replicating peptide. *Nature* 382, 525–528.

Levy, M. and Miller, S. (1998) The stability of the RNA bases: implications for the origin of life. *Proc. Natl. Acad. Sci. USA* 95, 7933–7938.

Li, A. (2005) Can fluffy dust alleviate the subsolar interstellar abundance problem? *Astrophys. J.* 622, 965–969.

Lindgren, K. (1992) Evolutionary phenomena in simple dynamics. In *Santa Fe Institute Studies in the Sciences of Complexity, Vol. X: Artificial Life II*, edited by C. Langton, C. Taylor, J.D. Farmer, and S. Rasmussen, Addison-Wesley, Redwood City, CA, pp. 295–312.

Lindgren, K. and Nordahl, M.G. (1994) Evolutionary dynamics of spatial games. *Physica D* 75, 292–309.

Linski, R.E., Ofria, C., Collier, T.C., and Adami, C. (1999) The minimal systems requirements for open-ended evolution is key open research question for evolutionary dynamics. *Nature* 400, 661–664.

Lohrmann, R. and Orgel, L. (1971) Urea-inorganic phosphate mixtures as prebiotic phosphorylating agents. *Science* 171, 490–494.

Luisi, P.L., Giomini, M., Pileni, M., and Robinson, B. (1988) Reverse micelles as hosts for proteins and small molecules. *Biochim. Biophys. Acta* 947, 209–246.

Luisi, P.L., Walde, P., and Oberholzer, T. (1994) Enzymatic synthesis in self-reproducing vesicles: an approach to the construction of a minimal cell. *Ber. Bunsenges. Phys, Chem*. 98, 1160–1165.

Luisi, P.L., Veronese, A., and Berclaz, N. (1998) Photoinduced formation of bilayer vesicles. *J. Phys. Chem. B* 102, 7078–7080.

Maden, B. (1995) No soup for starters? Autotrophy and origins of metabolism. *Trends Biochem. Sci*. 20, 337–341.

Mannings, V., Boss, A., and Russell, S.S. (2000) *Protostars and Planets IV*, University of Arizona Press, Tucson.

Marcy, G.W., Butler, R.P., Fischer, D.A., Vogt, S.S., Wright, J.T., and Tinney, C.G. (2005) Observed properties of exoplanets: masses, orbits, and metallicities. *Prog. Theor. Phys. Suppl*. 158, 24–42.

Markwick, A. and Charnley, S.B. (2004). Chemistry of protoplanetary disks: relation to primitive solar system material. In *Astrophysics and Space Science Library, Vol. 305: Astrobiology: Future Perspectives*, edited by P. Ehrenfreund, W.M. Irvine, T. Owen, L. Becker, J. Blank, J.R. Brucato, L. Colangeli, S. Derenne, A. Dutrey, D. Despois, A. Lazcano, and F. Robert, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 33–66.

Mathis, J.S., Mezger, P.G., and Panagia, N. (1983) Interstellar radiation field and dust temperatures in the diffuse interstellar matter and in giant molecular clouds. *Astron. Astrophys*. 128, 212–229.

McCollom, T., Ritter, G. and Simoneit, B. (1999) Lipid synthesis under hydrothermal conditions by Fischer-Tropsch-type reactions. *Orig. Life Evol. Biosph*. 29, 153–166.

Mennella, V., Colangeli, L., Bussoletti, E., Palumbo, P., and Rotundi, A. (1998) A new approach to the puzzle of the ultraviolet interstellar extinction bump. *Astrophys. J.* 507, 177–180.

Miller, S. (1953) A production of amino acids under possible primitive Earth conditions. *Science* 117, 528–529.

Miller, S. (1957) The mechanism of synthesis of amino acids by electric discharges. *Biochim. Biophys. Acta* 23, 480–489.

Miller, S. (1998) The endogenous synthesis of organic compounds. In *The Molecular Origins of Life. Assembling Pieces of the Puzzle*, edited by A. Brack, Cambridge University Press, Cambridge, UK, pp. 59–85.

Miller, S. and Bada, J. (1988) Submarine hot springs and the origin of life. *Nature* 334, 609–611.

Miller, S. and Orgel, L. (1974) *The Origins of Life on the Earth*, Prentice-Hall, Englewood Cliffs, NJ.

Miyakawa, S., Yamanashi, H., Kobayashi, K., Cleaves, H., and Miller, S. (2002a) Prebiotic synthesis from CO atmospheres: implications for the origins of life. *Proc. Natl. Acad. Sci. USA* 99, 14628–14631.

Miyakawa, S., Cleaves, H., and Miller, S. (2002b) The cold origin of life: A. Implications based on the hydrolytic stabilities of hydrogen cyanide and formamide. *Orig. Life Evol. Biosph*. 32, 195–208.

Miyakawa, S., Cleaves, H., and Miller, S. (2002c) The cold origin of life: B. Implications based on pyrimidines and purines produced from frozen ammonium cyanide solutions. *Orig. Life Evol. Biosph*. 32, 209–218.

Mojzsis, S., Arrhenius, G., McKeegan, K., Harrison, T., Nutman, A., and Friend, C. (1996) Evidence for life on Earth before 3,800 million years ago. *Nature* 384, 55–59.

Mojzsis, S., Harrison, T., and Pidgeon, R. (2001) Oxygen-isotope evidence from ancient zircons for liquid water at the Earth's surface 4,300 Myr ago. *Nature* 409, 178–181.

Monnard, P.A. and Deamer, D. (2002) Membrane self-assembly processes: steps toward the first cellular life. *Anat. Rec*. 268, 196–207.

Morowitz, H., Deamer, D., and Heinz, B. (1988) The chemical logic of a minimal protocell. *Orig. Life Evol. Biosph*. 18, 281–287.

Mosqueira, F., Albarran, G., and Negron-Mendoza, A. (1996) A review of conditions affecting the radiolysis due to 40K on nucleic acid bases and their derivatives adsorbed on clay minerals: implications in prebiotic chemistry. *Orig. Life Evol. Biosph*. 26, 75–94.

Muñoz Caro, G.M., Meierhenrich, U.J., Schutte, W.A., Barbier, B., Arcones Segovia, A., Rosenbauer, H., Thiemann, W.H.-P., Brack, A., and Greenberg, J.M. (2002) Amino acids from ultraviolet irradiation of interstellar ice analogues. *Nature* 416, 403–406.

Munteanu, A. and Sole, R. (2005) Chaos in chemoton dynamics. In *Santa Fe Institute Preprint Series*, Santa Fe Institute, Santa Fe, NM, 2005-05-017.

Nelson, K., Levy, M., and Miller, S. (2000) Peptide nucleic acids rather than RNA may have been the first genetic molecule. *Proc. Natl. Acad. Sci. USA* 97, 3868–3871.

Nielsen, P., Egholm, M., Berg, R., and Buchardt, O. (1991) Sequence-selective recognition of DNA by strand displacement with a thymine-substituted polyamide. *Science* 254, 1497–1500.

Nisbet, E. and Sleep, N. (2001) The habitat and nature of early life. *Nature* 409, 1083–1090.

Nuth, J.A. (1985) Meteoritic evidence that graphite is rare in the interstellar medium. *Nature* 318, 166–168.

Orgel, L. (1998) The origin of life—a review of facts and speculations. *Trends Biochem. Sci*. 23, 491–495.

Oró, J. (1960) Synthesis of adenine from ammonium cyanide. *Biochem. Biophys. Res. Commun*. 2, 407–412.

Oró, J. and Kimball, A. P. (1961) Synthesis of purines under primitive Earth conditions. I. Adenine from hydrogen cyanide. *Arch. Biochem. Biophys*. 94, 221–227.

Oró, J. and Lazcano, A. (1997) Comets and the origin and evolution of life. In *Comets and the Origin and Evolution of Life*, edited by P.J. Thomas, C.F. Chyba, and C.P. McKay, Springer, New York, pp. 3–27.

Osterberg, R. and Orgel, L. (1972) Polyphosphate and trimetaphosphate formation under potentially prebiotic conditions. *J. Mol. Evol*. 1, 241–248.

Pace, N.R. (2001) The universal nature of biochemistry. *Proc. Natl. Acad. Sci. USA* 98, 808–808.

Pasek, M. and Lauretta, D. (2005) Aqueous corrosion of phosphide minerals from iron meteorites: a highly reactive source of prebiotic phosphorus on the surface of the early Earth. *Astrobiology* 5, 515–535.

Peeters, Z., Botta, O., Ruiterkamp, R., Charnley, S.B., and Ehrenfreund, P. (2003) The astrobiology of nucleobases. *Astrophys. J*. 593, 129–132.

Peeters, Z., Botta, O., Charnley, S.B., Kuan, Y., Kisiel, Z., and Ehrenfreund, P. (2005) Formation and photostability of N-heterocycles in space: the effect of nitrogen on the photostability of small aromatic molecules. *Astron. Astrophys*. 433, 583–590.

Pendleton, Y. and Allamandola, L. (2002) The organic refractory material in the diffuse interstellar medium: mid-infrared spectroscopic constraints. *Astrophys. J. Suppl*. 138, 75–98.

Pinto, J., Gladstone, G., and Yung, Y. (1980) Photochemical production of formaldehyde in Earth's primitive atmosphere. *Science* 210, 183–185.

Pohorille, A. and Deamer, D. (2002) Artificial cells: prospects for biotechnology. *Trends Biotechnol*. 20, 123–170.

Prasad, S.S. and Tarafdar, S.P. (1983) UV radiation field inside dense clouds—its possible existence and chemical implications. *Astrophys. J*. 267, 603–609.

Rabinowitz, J. and Hampai, A. (1978) Influence of imidazole and hydrocyanic acid derivatives on the 'possible prebiotic' polyphosphate-induced peptide synthesis in aqueous solution. *Helv. Chim. Acta* 61, 1842–1847.

Rasmussen, S. (1988) Toward a quantitative theory of the origin of life. In *Artificial Life*, edited by C. Langton, Addison-Wesley, Redwood CA, pp. 79–104.

Rasmussen, S., Chen, L., Nilsson, M., and Abe, S. (2003) Bridging nonliving and living matter. *Artif. Life* 9, 269–316.

Rasmussen, S., Chen, L., Deamer, D., Krakauer, D., Packard, N., Stadler, P., and Bedau, M. (2004) Transitions from non-living to living matter. *Science* 303, 963–965.

Ray, T.S. (1991) An approach to the synthesis of life. In *Artificial Life II*, edited by C. Langton, C. Taylor, J.D. Farmer, and S. Rasmussen, Addison-Wesley, Redwood, CA, pp. 371–408.

Reid, C. and Orgel, L. (1967) Synthesis of sugar in potentially prebiotic conditions. *Nature* 216, 455–455.

Rietmeijer, F.J.M., Rotundi, A., and Heymann, D. (2004) C_{60} and giant fullerenes in soot condensed in vapors with variable C/H_2 ratio. *Fullerenes Nanotubes Carbon Nanostruct*. 12, 659–680.

Robertson, M. and Miller, S. (1995) An efficient prebiotic synthesis of cytosine and uracil. *Nature* 375, 772–774.

Rocheleau, T., Rasmussen, S., Nielsen, P., Nilsson Jacobi, M., and Ziock, H. (2006) *LANL Technical Report: Emergence of Protocellular Growth Laws*, LA-UR-05-8916, Los Alamos National Laboratory Technical Report. *Proc. R. Soc. Lond. B Biol. Sci.* (in press).

Rode, B. (1999) Peptides and the origin of life. *Peptides* 20, 773–786.

Ruiterkamp, R., Peeters, Z., Moore, M., Hudson, R. and Ehrenfreund, P. (2005a) A quantitative study of proton irradiation and UV photolysis of benzene in interstellar environments. *Astron. Astrophys*. 440, 391–402.

Ruiterkamp, R., Cox, N.L.J., Spaans, M., Kaper, L., Foing, B.H., Salama, F., and Ehrenfreund, P. (2005b) PAH charge state distribution and DIB carriers: implications from the line of sight toward HD 147889. *Astron. Astrophys*. 431, 515–529.

Russell, M. and Hall, A.J. (1997) The emergence of life from iron monosulphide bubbles at a submarine hydrothermal redox and pH front. *J. Geol. Soc. Lond*. 154, 377–402.

Russell, M.J. and Martin, W. (2004) The rocky roots of the acetyl coenzyme-A pathway. *Trends Biochem. Sci*. 24, 358–363.

Salama, F., Bakes, E.L.O., Allamandola, L.J., and Tielens, A.G.G.M. (1996) Assessment of the polycyclic aromatic hydrocarbon—diffuse interstellar band proposal. *Astron. Astrophys*. 458, 621–636.

Sanchez, R. and Orgel, L. (1970) Prebiotic synthesis. V. Synthesis and photoanomerization of pyrimidine nucleosides. *J. Mol. Biol*. 47, 531–543.

Sanchez, R., Ferris, J., and Orgel, L. (1966a) Conditions for purine synthesis: did prebiotic synthesis occur at low temperatures? *Science* 153, 72–73.

Sanchez, R., Ferris, J., and Orgel, L. (1966b) Cyanoacetylene in prebiotic synthesis. *Science* 154, 784–785.

Sanchez, R., Ferris, J., and Orgel, L. (1968) Studies in prebiotic synthesis. IV. The conversion of 4-aminoimidazole-5-carbonitrile derivatives to purines. *J. Mol. Evol*. 38, 121–128.

Sayama, H. (1998) Introduction of structural dissolution into Langton's self-replicating loops. In *Artificial Life VI: Proceedings of the Sixth International Conference on Artificial Life (Complex Adaptive Systems)*, edited by C. Adami, R.K. Belew, H. Kitano, and C.E. Taylor, MIT Press, Cambridge, MA, pp. 114–122.

Scalo, J. and Biswas, A. (2002) Turbulent compressibility of protogalactic gas. *Monthly Notices R. Astron. Soc*. 332, 769–776.

Scannapieco, E., Schneider, R., and Ferrara, A. (2003) The detectability of the first stars and their cluster enrichment signatures. *Astrophys. J.* 589, 35–52.

Schlesinger, G. and Miller, S. (1973) Equilibrium and kinetics of glyconitrile formation in aqueous solution. *J. Am. Chem. Soc.* 95, 3729–3735.

Schlesinger, G. and Miller, S. (1983) Prebiotic synthesis in atmospheres containing CH_4, CO, and CO_2. II. Hydrogen cyanide, formaldehyde and ammonia. *J. Mol. Evol*. 19, 383–390.

Schwartz, A. (1969) Specific phosphorylation of the 2′- and 3′-positions in ribonucleosides. *J. Chem. Soc. D Chem. Commun*. 23, 1393.

Segré, D. and Lancet, D. (2000) Composing life. *EMBO Rep*. 31, 217–222.

Segré, D., Ben-Eli, D., and Lancet, D. (2000) Compositional genomes: prebiotic information transfer in mutually catalytic noncovalent assemblies. *Proc. Natl. Acad. Sci. USA* 97, 4112–4117.

Segré, D., Ben-Eli, D., Deamer, D.W., and Lancet, D. (2001) The lipid world. *Orig. Life Evol. Biosphys*. 31, 119–145.

Sephton, M.A. (2002) Organic compounds in carbonaceous meteorites. *Nat. Prod. Rep*. 19, 292–311.

Sephton, M.A. and Botta, O. (2005) Recognizing life in the Solar System: guidance from meteoritic organic matter. *Int. J. Astrobiol.* 4, 269–276.

Sephton, M.A., Pillinger, C.T., and Gilmour I. (2000) Aromatic moieties in meteoritic macromolecular materials: analyses by hydrous pyrolysis and ^{13}C of individual compounds. *Geochim. Cosmochim. Acta* 64, 321–328.

Sephton, M.A., Verchovsky, A.B., and Wright, I.P. (2004) Carbon and nitrogen isotope ratios in meteoritic organic matter: indicators of alteration processes on the parent asteroid. *Int. J. Astrobiol*. 3, 221–227.

Shapiro, R. (1995) The prebiotic role of adenine: a critical analysis. *Orig. Life Evol. Biosph*. 25, 83–98.

Shock, E. and Schulte, M. (1998) Organic synthesis during fluid mixing in hydrothermal systems. *J. Geophys. Res. Planets* 103, 28513–28527.

Shock, E.L., McCollom, T., and Schulte, M.D. (1998) The emergence of metabolism from within hydrothermal systems. In *Thermophiles: The Keys to Molecular Evolution and the Origin of Life*, edited by J. Wiegel and M.W.W Adams, Taylor and Francis, Washington, pp. 59–76.

Simionescu, B., Mora, R., Leanca, M., and Ananiescu, C. (1980) Genesis of porphyrin-like compounds under simulated abiotic conditions. A possible mechanism of their formation. *Rev. Roum. Chim*. 25, 1077–1082.

Skusa, A. and Bedau, M.A. (2002) Towards a comparison of evolutionary creativity in biological and cultural evolution. In *Artificial Life VIII*, edited by R. Standish, M.A. Bedau, and H.A. Abbass, MIT Press, Cambridge, MA, pp. 233–242.

Sleep, N., Meibom, A., Fridriksson, T., Coleman, R., and Bird, D. (2004) H_2-rich fluids from serpentinization: geochemical and biotic implications. *Proc. Natl. Acad. Sci. USA* 101, 12818–12823.

Smith, E. and Morowitz, H. (2004) Universality in intermediary metabolism. *Proc. Natl. Acad. Sci. USA* 101, 13168–13173.

Snider, M. and Wolfenden, R. (2000) The rate of spontaneous decarboxylation of amino acids. *J. Am. Chem. Soc*. 122, 11507–11508.

Sollerman, J., Cox, N., Mattila, S., Ehrenfreund, P., Kaper,

L., Leibundgut, B., and Lundqvist, P. (2005) Diffuse interstellar bands in NGC 1448. *Astron. Astrophys*. 429, 559–567.

Sowerby, S., Morth, C., and Holm, N. (2001) Effect of temperature on the adsorption of adenine. *Astrobiology* 1, 481–487.

Spaans, M. (2004) The synthesis of the elements and the formation of stars. In *Astrophysics and Space Science Library, Vol. 305: Astrobiology: Future Perspectives*, edited by P. Ehrenfreund, W.M. Irvine, T. Owen, L. Becker, J. Blank, J.R. Brucato, L. Colangeli, S. Derenne, A. Dutrey, D. Despois, A. Lazcano, and F. Robert, Kluwer Academic Publishers, Dordrecht, The Netherlands, pp. 1–16.

Spaans, M. and Silk, J. (2005) The polytropic equation of state of primordial gas clouds. *Astrophys. J.* 626, 644–648.

Spoon, H.W.W., Moorwood, A.F.M., Pontoppidan, K.M., Cami, J., Kregel, M., Lutz, D., and Tielens, A.G.G.M. (2003) Detection of strongly processed ice in the central starburst of NGC 4945. *Astron. Astrophys*. 402, 499–507.

Stephan, B. (2002) The chemical basis of membrane bioenergetics. *J.Mol. Evol*. 54, 595–613.

Stillwell, W. (1977) On the origin of photophosphorylation. *J. Theor. Biol*. 65, 479–497.

Stribling, R. and Miller, S. (1987) Energy yields for hydrogen cyanide and formaldehyde syntheses: the hydrogen cyanide and amino acid concentrations in the primitive ocean. *Orig. Life Evol. Biosph*. 17, 261–273.

Szostak, J., Bartel, D., and Luisi, P. (2001) Synthesizing life. *Nature* 409, 387–390.

Taylor, R. and Walton, D.R.M. (1993) The chemistry of fullerenes. *Nature* 363, 685–693.

Tian, F., Toon, O., Pavlov, A., and De Sterck, H. (2005) A hydrogen-rich early Earth atmosphere. *Science* 308, 1014–1017.

Tielens, A.G.G.M., Hony, S., van Kerckhoven, C., and Peeters, E. (1999) Interstellar and circumstellar PAHs. In *ESA SP 427: The Universe as Seen by ISO*, edited by P. Cox and M.F. Kessler, European Space Agency, Noordwijk, The Netherlands, pp. 579–588.

Tomita, S., Fujii, M., and Hayashi, S. (2004) Defective carbon onions in interstellar space as the origin of the optical extinction bump at 217.5 nanometers. *Astrophys. J.* 609, 220–224.

Tuck, A. (2002) The role of atmospheric aerosols in the origins of life. *Surveys Geophys*. 23, 379–409.

Umeda, H. and Nomoto, K. (2005) Variations in the abundance pattern of extremely metal-poor stars and nucleosynthesis in population III supernovae. *Astrophys. J.* 619, 427–445.

van Zuilen, M.A., Lepland, A., and Arrhenius G. (2002) Reassessing the evidence for the earliest traces of life. *Nature* 418, 627–630.

Verlander, M., Lohrmann, R., and Orgel, L. (1973) Catalysts for the self-polymerization of adenosine cyclic 2′,3′-phosphate. *J. Mol. Evol*. 2, 303–316.

von Kiedrowski, G., Eckardt, L., Naumann, K., Pankau, W.M., Reimold, M., and Rein, M. (2003) Toward replicatable, multifunctional, nanoscaffolded machines. A chemical manifesto. *Pure Appl. Chem*. 75, 609–619.

von Neumann, J. (1966) *The Theory of Self Reproducing Automata*, edited by A.W. Burks, University of Illinois Press, Urbana.

Wächtershäuser, G. (1988) Before enzymes and templates: theory of surface metabolism. *Microbiol. Rev*. 52, 452–484.

Wächtershäuser, G. (1997) The origin of life and its methodological challenges. *J. Theor. Biol*. 187, 483–494.

Weidenschilling, S. and Cuzzi, J. (1993) Formation of planetesimals in the solar nebula. In *Protostars and Planets III*, edited by E. Levy and J. Lunine, University of Arizona Press, Tucson, pp. 1031–1060.

White, R. (1984) Hydrolytic stability of biomolecules at high temperatures and its implication for life at 250°C. *Nature* 310, 430–432.

Wilde, S., Valley, J., Peck, W., and Graham, C. (2001) Evidence from detrital zircons for the existence of continental crust and oceans on the Earth 4.4 Gyr ago. *Nature* 409, 175–178.

Wooden, D., Charnley, S., and Ehrenfreund, P. (2004) Composition and evolution of molecular clouds. In *COMETS II*, edited by M. Festou, H.U. Keller, and H. Weaver, University of Arizona Press, Tucson, pp. 33–66.

Yamagata, Y., Watanabe, H., Saitoh, M., and Namba, T. (1991) Volcanic production of polyphosphates and its relevance to prebiotic evolution. *Nature* 352, 516–519.

Yan, L., Chary, R., Armus, L., Teplitz, H., Helou, G., Frayer, D., Fadda, D., Surace, J., and Choi, P. (2005) Spitzer detection of polycyclic aromatic hydrocarbon and silicate dust features in the mid-infrared spectra of $z \sim 2$ ultraluminous infrared galaxies. *Astrophys. J.* 628, 604–610.

Yuen, G. and Kvenvolden, K. (1973) Monocarboxylic acids in Murray and Murchison carbonaceous meteorites. *Nature* 246, 301–303.

Zhang, L., Peritz, A., and Meggers, E. (2005) A simple glycol nucleic acid. *J. Am. Chem. Soc*. 127, 4174–4175.

Zhaxybayea, O. and Gogarten, J.P. (2004) Cladogenesis, coalescence and the evolution of the three domains of life. *Trends Genet*. 20, 182–187.

Address reprint requests to:
Prof. Pascale Ehrenfreund
Astrobiology Laboratory
Leiden Institute of Chemistry
P.O. Box 9502
2300 RA Leiden, The Netherlands

E-mail: p.ehrenfreund@chem.leidenuniv.nl

http://www.astrobiology.nl

This article has been cited by:

1. Boncho P. Bonev, Michael J. Mumma, Erika L. Gibb, Michael A. DiSanti, Geronimo L. Villanueva, Karen Magee-Sauer, Richard S. Ellis. 2009. COMET C/2004 Q2 (MACHHOLZ): PARENT VOLATILES, A SEARCH FOR DEUTERATED METHANE, AND CONSTRAINT ON THE CH 4 SPIN TEMPERATURE. *The Astrophysical Journal* **699**:2, 1563-1572. [CrossRef]
2. Sun Kwok. 2009. Delivery of Complex Organic Compounds from Planetary Nebulae to the Solar System. *International Journal of Astrobiology* 1. [CrossRef]
3. V. C. Geers, E. F. van Dishoeck, K. M. Pontoppidan, F. Lahuis, A. Crapsi, C. P. Dullemond, G. A. Blake. 2009. Lack of PAH emission toward low-mass embedded young stellar objects. *Astronomy and Astrophysics* **495**:3, 837-846. [CrossRef]
4. César Menor-Salván, Marta Ruiz-Bermejo, Susana Osuna-Esteban, Guillermo Muñoz-Caro, Sabino Veintemillas-Verdaguer. 2009. Synthesis of Polycyclic Aromatic Hydrocarbons and Acetylene Polymers in Ice: A Prebiotic Scenario. *Chemistry & Biodiversity* **5**:12, 2729-2739. [CrossRef]
5. Andrew D. Aubrey , John H. Chalmers , Jeffrey L. Bada , Frank J. Grunthaner , Xenia Amashukeli , Peter Willis , Alison M. Skelley , Richard A. Mathies , Richard C. Quinn , Aaron P. Zent , Pascale Ehrenfreund , Ron Amundson , Daniel P. Glavin , Oliver Botta , Laurence Barron , Diana L. Blaney , Benton C. Clark , Max Coleman , Beda A. Hofmann , Jean-Luc Josset , Petra Rettberg , Sally Ride , François Robert , Mark A. Sephton , Albert Yen . 2008. The Urey Instrument: An Advanced In Situ Organic and Oxidant Detector for Mars ExplorationThe Urey Instrument: An Advanced In Situ Organic and Oxidant Detector for Mars Exploration. *Astrobiology* **8**:3, 583-595. [Abstract] [PDF] [PDF Plus]
6. J. L. Bada, P. Ehrenfreund, F. Grunthaner, D. Blaney, M. Coleman, A. Farrington, A. Yen, R. Mathies, R. Amudson, R. Quinn, A. Zent, S. Ride, L. Barron, O. Botta, B. Clark, D. Glavin, B. Hofmann, J. L. Josset, P. Rettberg, F. Robert, M. Sephton. 2008. Urey: Mars Organic and Oxidant Detector. *Space Science Reviews* **135**:1-4, 269-279. [CrossRef]
7. Clifford N. Matthews, Robert D. Minard. 2008. Hydrogen cyanide polymers connect cosmochemistry and biochemistry. *Proceedings of the International Astronomical Union* **4**:S251, 453. [CrossRef]
8. V.N. Kompanichenko. 2008. Three stages of the origin of life process: bifurcation, stabilization and inversion. *International Journal of Astrobiology* **7**:01. . [CrossRef]
9. John Parnell , David Cullen , Mark R. Sims , Stephen Bowden , Charles S. Cockell , Richard Court , Pascale Ehrenfreund , Francois Gaubert , William Grant , Victor Parro , Michel Rohmer , Mark Sephton , Helga Stan-Lotter , Andrew Steele , Jan Toporski , Jorge Vago . 2007. Searching for Life on Mars: Selection of Molecular Targets for ESA's Aurora ExoMars MissionSearching for Life on Mars: Selection of Molecular Targets for ESA's Aurora ExoMars Mission. *Astrobiology* **7**:4, 578-604. [Abstract] [PDF] [PDF Plus]
10. Pascale Ehrenfreund, Mark A. Sephton. 2006. Carbon molecules in space: from astrochemistry to astrobiology. *Faraday Discussions* **133**, 277. [CrossRef]
11. A. G. Donchev, N. G. Galkin, L. B. Pereyaslavets, V. I. Tarasov. 2006. Quantum mechanical polarizable force field (QMPFF3): Refinement and validation of the dispersion interaction for aromatic carbon. *The Journal of Chemical Physics* **125**:24, 244107. [CrossRef]

ComPlexUs
BIO INSPIRED METHODS
Complexus 2006;3:48–57 DOI: 10.1159/000094187
Published online: August 25, 2006

An Electronically Controlled Microfluidic Approach towards Artificial Cells

Uwe Tangen[a] Patrick F. Wagler[a] Steffen Chemnitz[a]
Goran Goranovic[b] Thomas Maeke[a]
John S. McCaskill[a, c]

[a]Ruhr-Universität-Bochum, Biomolecular Information Processing (BioMIP), Dortmund, Germany
[b]MEMPHYS Center for Biomembrane Physics, Physics Department, University of Southern Denmark, Odense, Denmark;
[c]Friedrich-Schiller-Universität, Jena, Germany

Key Words
Reconfigurable hardware · Microfluidics · Micro-reactor · Polydimethylsiloxane · DNA · Microelectrode · Fluorescence detection

Abstract
This work focuses on the application of on-line programmable microfluidic bioprocessing as a complementation vehicle towards the design of artificial cells. The electronically controlled collection, separation and channel transfer of the biomolecules are monitored by a sensitive fluorescence setup. Two different physical effects, electrophoresis and electroosmotic flow, are used to allow for a detailed micro-control of fluids in micro-reaction environments. A combination of these two basic electronically controlled input reaction chambers makes com-

Fax +41 61 306 12 34
E-Mail karger@karger.ch
www.karger.com
KARGER

1424–8492/06/0033–0048
$23.50/0
Accessible online at:
www.karger.com/cpu

Dr. Uwe Tangen, Ruhr-Universität-Bochum, Biomolecular Information Processing (BioMIP), c/o BMZ, Otto-Hahn-Str. 15
DE-4427 Dortmund (Germany)
Tel. +49 231 9742 6424, Fax +49 231 9742 6426
E-Mail uwe.tangen@biomip.ruhr-uni-bochum.de

Simplexus

While François Jacob, one of the pioneers in the study of genetic regulatory mechanisms, spoke about the 'logic of life' at the end of the 1960s, it is only in the past decade or so that a picture has begun to emerge of how the mechanisms of cell biology can be related to the concepts developed in the information sciences. This synthesis is represented by the new discipline of systems biology, in which the living cell is considered as a network of interacting processes that are coordinated and regulated according to some of the same principles used by engineers to control complex artificial systems, most notably those embodied in (micro)electronic circuitry. Put simply, systems biologists (a community that draws on the skills of biologists, physicists, engineers and mathematicians) regard the cell as something like a 'biological circuit board' that orchestrates its modules and functions to achieve robust, reliable and predictable operation.

To some extent this perspective was embedded already in the way that metabolic processes have long been depicted as complex networks of biochemical, generally enzyme-catalyzed, reactions. But in systems biology these schemes are regarded more explicitly as flows of information, involving operations such as feedback, synchronization, amplification and error correction that are familiar to engineers. Moreover, there is a crucial element of 'insulation' between the various components of the network, achieved in the cell partly by spatial compartmentalization and partly by chemical specificity (molecular recognition) among the biomolecular constituents. The cell is now seen as much more than a mere bag of chemicals.

The level of abstraction inherent in systems biology – in which, for example, a metabolic operation can be depicted as a 'circuit diagram' that clearly bears no visual resemblance to what is seen of a cell under the microscope – suggests that attempts to mimic cell functions, or even to

binatorial fluidic networks and indefinitely sustained biochemical or chemical reaction networks feasible. Experimental data showing the power of this approach is presented. Not only does this processing power pave the way towards the development of artificial cells (using a technology to complement not yet established autonomous metabolic or replication capabilities) but it also opens up new processes for applications of combinatorial chemistry and lab-on-a-chip biotechnology to drug discovery and diagnosis.

Introduction

Integrated microfluidics holds the promise of allowing complex parallel and pipelined processing of combinatorial families of biomolecules without the serial limitations of external robotic pipetting [1]. They allow one to build up dynamic micro-reactors and biochemical networks that are user-programmable, in space and time. Through operational feedback and multiple reconfiguration, they can also be iteratively optimized in an evolutionary process [2]. The development of integrated applications in biotechnology is limited by the necessity of custom hardware and time-consuming development cycles [3–6]. Reconfigurability opens new possibilities, as one patented strategy allowing researchers to proceed from simple connected micro-reactor elements to dynamically reconfigurable micro-reactor networks [7] shows. These consist of a combinatorial array of processing and switching elements, which can be externally programmed.

Examples of applications are microscale reactors for evolutionary biotechnology, active microfluidics for μTAS and molecular computing in micro-reactors [8–10]. Artificial containers such as droplets and vesicles in microfluidic environments represent an important area of research in down-scaling combinatorial chemistry but we have proposed that they can also assist to engineer artificial [11, 12] and minimal [13] cells. The important idea in this respect is to use the electronically controlled environment as a substitute for controlled cell functions like metabolism and/or replication that cannot yet be autonomously regulated. We analyzed the integration of digital microfluidics using immiscible fluids in closed channels, for droplet generation, fusion and transport. Reactors were fabricated in micro-molded PDMS (polydimethylsiloxane) or bonded silicon to glass. We extended previous work [14] to integrate continuous droplet generation and autonomously regulated serial droplet fusion [15]. In the experiments reported here, we incorporated in our microfluidic technology electroosmotic flow (EOF) [16] in addition to electrophoresis and show its feasibility.

The main unit of our system is an electronically programmable microfluidic chip equipped with micro-electrodes and fluidic channels and chambers, linked to a reconfigurable electronic chip. The microelectrodes are the actuator components in the microfluidic network (see fig. 1) for the EOF reaction chamber used in the experiments reported here. Furthermore, the chip contains a standard field-programmable gate array (FPGA). The reconfigurable logic device is integrated to control a maximum number of electrodes individually. With this intimate control of biochemical processes a precondition for 'live control' comes within reach, namely that the seamless integration of computer technology and biology occurs in both directions simultaneously. This involves both the real time detection and data analysis of the biomolecular system and the active control of biomolecules in a hybrid system [17]. Currently most of our experiments use small DNA oligomers (21 nt, Rh6G-GATGGTCACAGCATGTCTGTA) to test the setup.

Three physical effects are apparent using this technology:

build entire 'artificial cells', need not slavishly copy the cell's hardware. Just as there is a universality to computing (implied by the early work of Alan Turing), so that identical computations can be performed by banks of transistors on a microchip and by balls shuttling down arrays of interconnected tubes, so it is reasonable to think of an 'artificial cell' made from networks linking chemical reservoirs, with computer-controlled transfer of reagents between them. Uwe Tangen and colleagues describe a first step towards developing this kind of 'cell on a chip'.

The creation of the network for a complete artificial cell is far too daunting a challenge at this stage of the game. But that may not in any case be the most important objective. The ability to control the interactions of many different chemical reagents in time and space could be extremely valuable for biotechnology, perhaps providing an alternative to the current use of engineering microorganisms for biosynthesis of valuable biochemicals and pharmaceuticals. It can be extremely difficult to engineer complex, multi-stage synthetic pathways into real cells, in part because it is hard to alter one part of the cellular biochemical network without perturbing other parts. A synthetic 'bioreactor' should in principle offer easier control, for example because the 'housekeeping' pathways that keep the cellular wheels turning could be replaced by active, external control of the flows through the artificial network.

More dramatically, an adaptive control mechanism might enable an evolutionary aspect to be built into such an artificial network, so that it can be optimized for a particular outcome. Reproducing subsections of a full cellular network in an artificial system could allow the testing of the effects of new drug candidates on specific cell functions. The identification of such prospective drugs might itself be assisted by the kind of combinatorial chemistry that micro-networks permit. And the possibility of performing real computations using in-

Electrophoresis

The electrical field is quite inhomogeneous due to the small electrodes (e.g. 10–40 μm) and high field strengths. DNA contains non-aligned permanent dipole moments and a permanent negative charge from the phosphate groups which is partially offset by counterions from the solvent. Two forces cause electrophoretic molecular transport. The first one is the force of the field on the net charge. This is the same force that drives the ionic constituents of an electrolyte through the solution. The second is the force on the dipole of the molecule: which increases with the divergence of the field. Besides permanent dipole moments DNA also has significant induced dipole moment forces. Note that the strong molecular length dependence of DNA in gels (amplified by chain entanglement) is much less pronounced in free solution. The electrophoretic mobility in free solution in the simplest analysis is the ratio of net charge to hydrodynamic resistance (which follows the Stokes law $F = 6\pi f\eta\, R$ with an asymmetry correction factor f).

Electroosmotic Flow

In translation invariant geometry (infinite channels) the Hagen-Poiseuille (H-P) law is valid for a steady-state pressure-driven flow of incompressible fluid

$$\triangle p \;=\; R_{\mathrm{hyd}} Q \qquad (1)$$

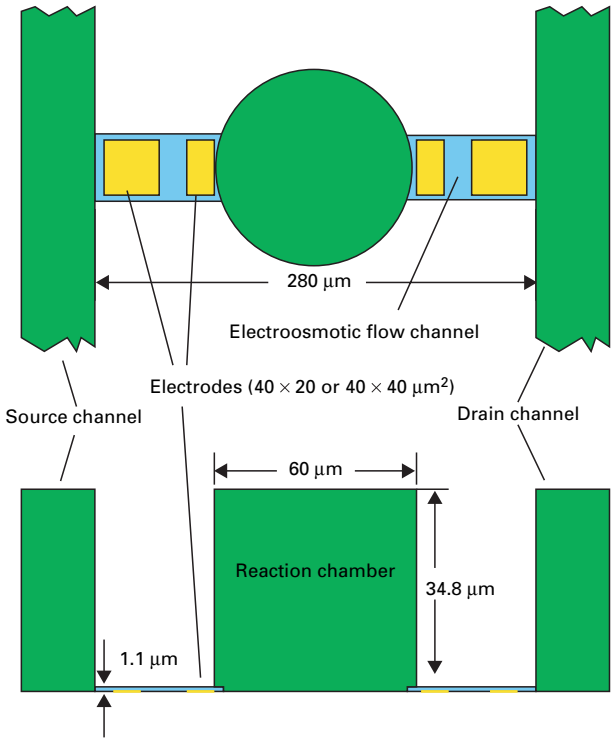


Fig. 1. Electronic-regulated chamber for an artificial cell connected to chemical supply and drain with two IO channels of 1.1 μm in height [measured with profiler P-10 (Fa. KLA-Tencor, USA) and White-Light-InterferometerWyko NT 1000 Optical Profiler (Veeco Process Metrology)] and regulated by two gold electrodes placed below each I/O. The reaction chamber, micro-moulded (SU-8 master) in PDMS with a size of 34.8 by 60 μm diameter, is effectively decoupled from the hydrodynamic flow of the supply and drain channels, while the programmable electrodes enable adjustable bulk EOF and electrophoresis. The entire structure can form a basis of a system of programmable microfluidic networks (e.g. fig. 6).

formation encoded in DNA molecules, proposed over a decade ago, might also become more feasible if the biochemical principles involved are integrated with the kind of logic operations realizable in micro-reactor networks.

Tangen et al. have so far focused on perhaps the most fundamental process that would need to be conducted in an artificial biochemical network: the controlled transport of a reagent into and out of a reaction chamber. These chambers are connected to one another, and to reservoirs of reagents, by narrow channels. Such components can be rendered at the microscopic scale by using standard microfabrication methods to carve molds into silicon wafers, and then using these to cast structures in polymers such as the elastomer polydimethylsiloxane (PDMS). In the experiments reported here, two flow channels about 40 μm wide and 1 μm high feed into a circular reaction chamber 60 μm in diameter and 34.8 μm high.

To create and control the fluid flow, the researchers place individually addressable metal electrodes on the base of the flow channels. The charging configurations used in the current prototype are simple, but in a more complex network the electrodes may be controlled 'on the fly' by a field-programmable gate array (a reconfigurable logic device), enabling active and continual reprogramming of the flow patterns.

The flow is driven by two key phenomena. In electrophoresis, either permanent or field-induced electric dipole moments on the solute molecules cause them to be attracted towards the charged electrodes. This is the same as the process that enables (polyanionic) DNA and other charged biomolecules to be separated according to their mobility in the standard analytical technique of gel electrophoresis. But flow is also induced, in these electrolyte solutions, by electroosmosis. Charges on the walls of the channels (which are generally negative both on silica glass surfaces and

which says that the flow rate Q is proportional to the pressure drop. Electroosmosis, on the other hand, is a non-equilibrium effect, where a liquid is brought to move relative to a charged surface an applied external electric field by acceleration of ions in the solution double layer which screens the surface charge. Surface double layers are characterized by the ζ-potential. For an ideal EOF at an infinite plate (uniform ζ-potential along the wall, homogeneous E field, steady-state flow, the Debye layer much thinner than the smallest channel dimension), the electroosmotic velocity is given by

$$u_{eo} = \alpha_{eo} E = \frac{\epsilon\zeta}{\mu} E. \quad (2)$$

Here α_{eo} is the electroosmotic mobility, ζ is the ζ-potential of the wall, ε and μ are the dielectric constant and dynamic viscosity of the electrolyte, respectively, and E is the applied electric field. Total EOF in a micro-channel is then the sum of individual EOFs from each wall, each approximated by equation 2.

In order to obtain stable fluid control into and out of the reaction chambers (fig. 1), it is crucial that pressure-driven hydrodynamic flow from the source and the drain channels [3] is suppressed. Flow control can be obtained despite high resistance using EOF in many small or a single high aspect ratio channel. To underline this essential mechanism, we first look at an isolated rectangular micro-channel of the length L, width W, and height h, with an (infinitely) thin double layer on all walls. The EOF rate in it is given by

$$Q_{eo} = u_{eo} h W = \alpha_{eo} V_{\text{eff}} \frac{hW}{L} \propto, \quad (3)$$

where u_{eo} is the electroosmotic velocity, α_{eo} the electroosmotic mobility, and V_{eff} the electric potential drop inside the channel. Note that this expression implies that electrodes at fixed voltage, placed at either end of a very short electroosmotic channel segment, would produce a very large (diverging) volume flow. Clearly the reason for the problem is that the hydrodynamic resistance of the remaining microfluidic system has not been included. For a high aspect ratio, i.e., $W \gg h$, the hydraulic resistance of the channel is

$$R_{\text{hyd}} = \frac{12\mu L}{h^3 W} \frac{1}{1 - 0.63\frac{h}{W}}, \quad (4)$$

where μ is the dynamic viscosity. The pressure-driven flow rate Q_p through the channel is given by equation 1 with $Q_p \propto h^3$. From equations 3 and 1 it seems that pressure-driven flow will be negligible compared to the EOF for small values of h, but this only applies to loadless channels.

In the case of electroosmotic channels with a load, e.g. connected to a field-free channel segment of hydrodynamic resistance R_L, the lowest order approximation is to use Kirchoffs laws in connection with the low Reynolds number (creeping) flow. Basically, the volume flow for the loadless electroosmotic-driven segment is the resting state, and reductions of this flow velocity by the load cause a linear pressure difference across the electroosmotic segment, which can be calculated by equation 1. Kirchoff's laws imply that the pumped flow velocity satisfies

$$\triangle p_{\text{ext}} + \triangle p_{eo} + \triangle p_L = 0 \quad (5)$$

with the volume flows in the osmotic flow and load channels equated

$$Q_L = \frac{\triangle p_L}{R_L} = Q_{eo} + \frac{\triangle p_{eo}}{R_{eo}} \quad (6)$$

The hydrodynamic resistance is calculated from equation 4 and the loadless EOF from equation 3. Rearranging, we find

$$Q_L = \frac{R_{eo} Q_{eo} - \triangle p_{\text{ext}}}{R_{eo} + R_L} \quad (7)$$

which for zero external pressure difference reveals a reduction in the loadless osmotic flow rate by a factor of

$$r = \frac{R_{eo}}{R_{eo} + R_L}.$$

The underlying additivity assumption is based on a sharp transition from a plug flow profile to a parabolic flow profile at the junction of the driving and load channels. on the PDMS medium used here) create a so-called double-layer – a predominance of counterions close to the walls. An electric potential gradient in the direction of the channel, due to the electrodes, produces a force on these counterions that draws them towards the counter-electrode; and because these ions are solvated, they pull the solvent with them and create net fluid flow.

Using fluorescently labelled single-stranded DNA oligomers (21 bases) as the trial reagent, Tangen et al. show that they can use electrode charging to draw fluid from a source channel into the reaction chamber, and to drain it again. When an electrode on the 'far side' of the chamber is positively charged, the chamber fills with DNA in 2 or 3 min. Passive outflow, with the electrodes switched off, is very slow; but by reversing the electrode charging, or by charging up the inflow channel so as to repel the solute through the outflow channel, the chamber can be emptied to background levels in under a minute.

To implement a kind of 'flow logic' in more complex microfluidic networks is considerably more challenging. In particular, this is complicated by the fact that, while in electronic circuitry the signals can be digital (pulsed), these chemical networks are inescapably continuous (analog) systems that are at the mercy of the stochastic nature of molecular motions and detection. The question of how to develop control strategies in combined digital-analog systems has been studied previously in the context of classical engineering, where it was shown that it is not always possible to determine unambiguously that a particular logic operation has taken place. One of the key requirements, Tangen and colleagues say, will be the implementation of detection thresholds for solute concentrations that essentially provide a digital readout from an analog signal. They confess that more experimental work is needed to explore the feasibility of such strategies, although they say that prelimi-

For our experimental setup, the dimensions of the electroosmotic segments are $W \times L \times h = 40 \times 40 \times 1.1\ \mu m^3$, and the total load resistance is dominated by the electroosmotically non-active part in the horizontal channel; the sink resistors are, compared to this, two orders of magnitude smaller and the source does not even contribute to the total load resistance. This leads to $r = 0.18$, implying a 6-fold reduction in the electroosmotic-induced velocity. For an overall electroosmotic mobility of 4×10^{-4} cm^2/Vs, an effective voltage of 3.3 V, the channel geometry as above, $Q_{eo} = 0.52$ µl/h and without an external pressure difference, the flow in the horizontal channel is given by $Q_L = 0.1$ µl/h. In our experiments the non-reversed EOF is directed towards the cathode, indicating negatively charged walls and a positively charged double layer. In PDMS, one common problem is the poorly defined EOF and its dependence on the process used to seal the chip. While it has been reported that an oxidation step is required to produce EOF [18] others claim that no such step is necessary [19]. The use of multiple materials in a single device, each with a unique ζ-potential, may create non-uniform flows. Currently we cannot distinguish between the EOF created by the SiOxNy layer and that created by PDMS walls. The charged site on SiO_2 layers is invariably Si-OH, regardless of the silica type [20] and the presence of silicon nitride also results in predominantly Si-OH charges, with a minority of Si-NH_2 charges shifting the ζ-potential slightly toward positive values, so that the composed layer exhibits roughly similar results. Liu et al. [21] measured the electroosmotic mobility of native and oxidized PDMS/glass devices, both in a range of 4×10^{-4} cm^2/Vs at neutral pH, while, and in contrast to this, Lin et al. [22] investigated the temporary effect of the plasma modified PDMS and found that the ζ-potential decays to zero within 2 days. Finally, Bianchi et al. [23] presented a model allowing the determination of each individual ζ-potential in composite micro-channels. Here it remains for future work to clearly identify all electrophysical properties.

Electrolysis

The applied voltage of 3.3 V implies a highly positive reduction potential. Thus it would be possible to perform quite easily an oxidation of the DNA oligomers used in our experiments, since the redox potentials of DNA are in the range of a few hundred millivolts [24]. Furthermore, the applied potentials should lead to an electrolysis of water of which the redox potential is +1.23 V (O_2/H_2O). Although it is not easy to produce visible electrolysis in the buffer employed in our experiments, this will become an issue if electrodes are driven with DC for more than 10 s. Fortunately, it turned out that the experiments with EOF, even when statically driving the electrodes (at constant voltage) for 10 min did not show electrolysis. The major reason might be due to the extreme surface-volume ratio (1.1 µm height of the channels, see fig. 1). The EOF produced high fluid velocities inside the shallow channels and thus gas bubbles would be quickly transported away from the electrical field.

nary results from a five-input system, designed to enable combinatorial chemistry between these various ingredients, show promise in terms of the controllability and stability of the input flows. They suggest that the physical dimensions of the channels may constitute one of the important control parameters here, but a better understanding of the factors influencing the electroosmotic flow may also be needed.

Philip Ball
p.ball@nature.com

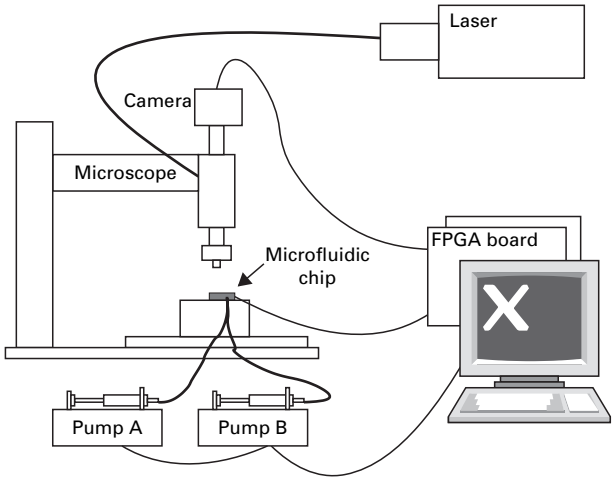


Fig. 2. Overall experimental setup. Camera Vosskuehler VDS-QCam 1300QLN (1280*1024, 11 fps), laser argon 514 nm, about 100 mW at the chip, MicroMechatronic Technologies AG (MMT) pumps (0.2–240 µl/h with 10 µl syringes), custom-designed and custom-built MereGen-FPGA board [25], Linux-based VME-bus host computer (Concurrent Technologies VP 100/01x Pentium III 1.2 Ghz, 512MB), microscope Zeiss Axiovert. The user-friendly control software is custom-designed.

Overall Experimental Setup

To distribute the DNA and buffer solutions on the chip a liquid-handling system was connected (see fig. 2). The microfluidic workstation includes high-performance syringe pumps for ultralow flow rates (MMT). The detection system is based on laser-induced fluorescence and a special optical system as well as a CCD camera for image acquisition. The electrode design was realized with a standard CAD schematic entry system for PCBs (Board Station Mentor Graphics). The special computer hardware MereGen™ based on reconfigurable electronics [25] analyzes the inner state of the micro-reactor network and realizes the programmability of the system using feedback loops.

The microfluidic device is equipped with gold electrodes (size typically 20 × 40 µm or 40 × 40 µm, 96 per module) and a standard FPGA (Spartan XL XCS20CSP144 Xilinx) that serves as the driver for the electrodes and as the interface to the MereGen board. The electrodes themselves are driven via the standard output drivers of the FPGA meaning a digital 3.3 V signal level. With the three states provided (3.3 V, 0 V and tristate) fast cycling allows us to emulate effectively arbitrary voltage levels between both extreme values. The configurable logic of the FPGA also serves as a testing device for the communication buses and built-in self-tests.

When molecules are moved, or chemical properties in the fluidic-system changed, fluorescence or white-light imaging using the camera mounted on the microscope gives the possibility of direct online control (see fig. 2). The camera image is preprocessed on the MereGen board. Arbitrary regions selected either automatically, via image processing, or manually may be evaluated, statistically processed and subjected to control rules. Depending on the complexity of these calculations, at least the hardest real-time critical components can be realized in hardware. Introducing low-level control structures paves the way to connect this work with computer science on hybrid systems.

A custom-developed software and operating system, with a user-friendly graphical interface, gives the experimenter easy access to the underlying hardware. A nice feature of this software is the remote-operating capability.

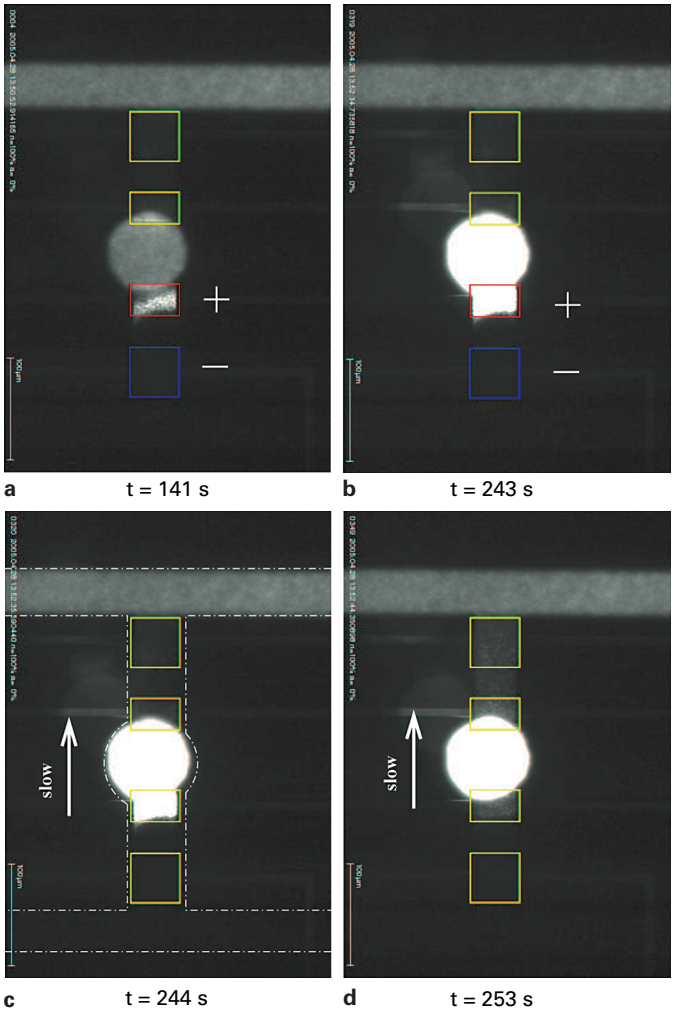


Fig. 3. Experimental demonstration of electronically programmable cell filling. Four samples at different times of a filling experiment are shown. The upper horizontal channel (depicted in **c** as dashed white lines) is filled with Rh6G $5 \cdot 10^{-6}$ labelled 21mer primer Rev500 batch No. 33436N in histidine buffer 50 mM, 59 µS/cm, pH 7.7 and excited with laser light at 514 nm with a power of 100 mW. At the beginning of the experiment no fluorescence could be observed in the reaction chamber. Two different physical effects contribute to filling the reaction chamber: electrophoresis (the positive electrode, 3.3 V, **a** and **b**) attracts the negatively charged DNA and electroosmotic flow which drags fluid from the upper supply channel against the weak background flow from the lower drain channel upwards. After releasing the electrodes (tristate, high impedance), material from the reaction chamber is slowly washed out again. Because of the considerable difference in transport capacity of the input/output channels and the storage capacity (see fig. 1) of the reaction chamber, this process is quite slow. With an appropriate presetting of the electrode potentials, even this slow dilution can be prevented.

Experimental Results

Two different solutions in two Hamilton syringe pumps were connected to the two inlets: one solution contained just buffer solution and the other rhodamine 6G-labelled single-stranded DNA with a length of 21 nt. A pump rate chosen between 0 and 2 µl/h was maintained stably in the supply and drain channels. The buffer solution employed was a freshly prepared histidine buffer (50 mM, 59 µS/cm, pH 7.7) to minimize the electrostatic screening effects (Debye length) [22].

The first task of the experiment was to fill the reaction chamber (see fig. 1) with the labelled DNA. The pump rates were adjusted such that a slow net flow from bottom to top (drain to supply) was maintained if no electrodes were active. At the beginning of the experiment, a faint fluorescence in the DNA supply ($5 \cdot 10^{-6}$ M) channel was visible. The filling procedure started with setting the second electrode from below to 3.3 V, and the bottom electrode to 0 V, as shown in figure 3a. The two field effects, EOF and electrophoresis, help each other in this case. The negatively charged DNA is attracted from the supply channel and in addition a net EOF is generated which pumps fluid from the supply channel towards the drain channel. On the other hand, the electrical field is polarized such that the DNA is repelled from the bottom electrode and attracted by the upper electrode. This results in the observed increase of the concentration of DNA (see fig. 3a–c). Note also the times involved in this filling procedure. Even after more than 200 s no electrolysis was observed. Immediately after releasing the electrodes, the slow volumetric flux (from bottom to top of the image) washes the material from the formerly positive electrode back into the reaction chamber (see fig. 3d). Despite the background flow, the time needed to wash out all the material from the reaction chamber is considerable as a result of the extreme height ratio between the reaction chamber and feeding channels. Further experiments showed (data not shown here) that, with suitably activated electrodes, even this weak outwash process can be drastically reduced and that additionally a reaction in the chamber might be fed continuously.

The second task was to drain the reaction chamber again. Two possible draining directions are available: pumping the material back into the supply channel (data not shown) which works remarkably fast (typically about 20 s are needed to reduce the fluorescence intensity to background level again), and pumping the material against the default flow pressure into the drain channel (see fig. 4). A time series of six images is presented. The experiment differed a little from the filling experiment in that the concentration of DNA in the supply channel was chosen 10 times

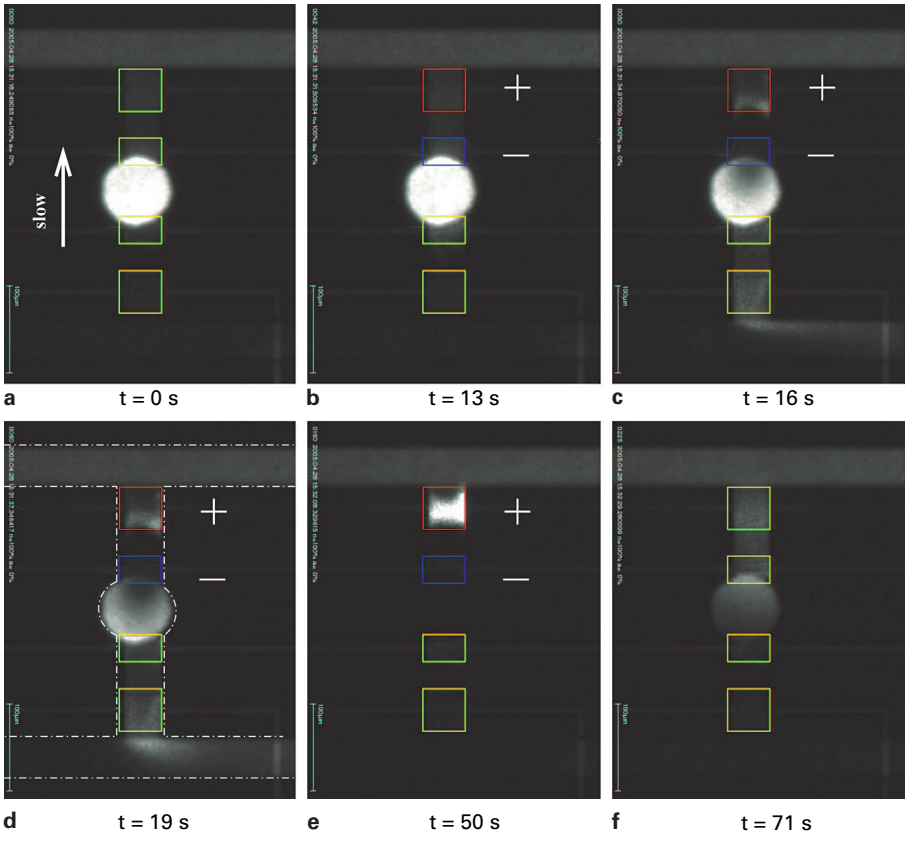


Fig. 4. Experimental demonstration of electronically programmable cell draining. Pumping material out of the reaction chamber is shown in the sequence of six photographs taken at different times. The starting situation is similar to the end situation shown in figure 3d. The concentration of the DNA in the supply chain has been reduced by a factor of ten ($5 \cdot 10^{-7}$ M) before filling the reaction chamber. The two physical forces, electrophoresis and electroosmosis, empty the reaction chamber (rapidly considering the huge size of the reaction chamber in comparison to the shallow input/output channels) in seconds. A side effect of electrostatic forces is the uptake of DNA material from the supply channel (upper channel, sketched as white dashed lines in **d**). This side effect can be used to create packets of charged material and transport this into a following channel system. As can be seen in figure 3, a slow default stream upwards cleans up the reaction chamber after the experiment. Additional suitable electrode control can drastically clean the reaction chamber (data not shown).

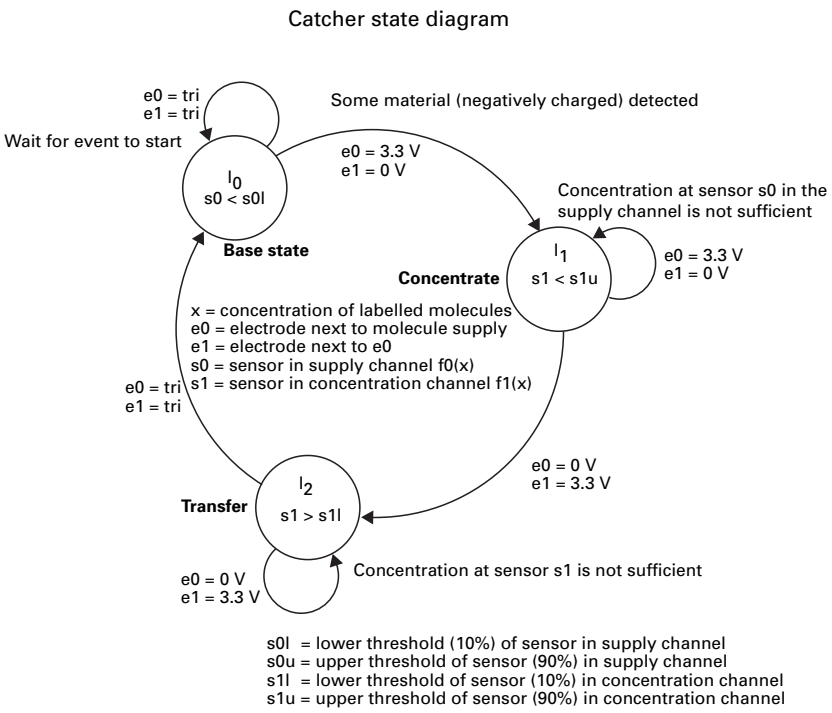


Fig. 5. A state machine diagram [notation used from [26] of one basic control module called 'the catcher'. Due to the on-line detection system actively controlled procedures are possible. This one serves for attracting charged material and generating an event to a following stage of the control network. Such control loops form the link to bridge the gap between microfluidic experimentation and global programming of biochemical processes with software. Noise, detection problems and other real-world effects can thus be reduced.

smaller, $5 \cdot 10^{-7}$ M. Everything else remained the same. In this case, the top electrodes are operated, reversing the EOF and expelling the DNA from the reaction chamber into the drain channel. After about 50 s, the fluorescence intensity in the reaction chamber was again indistinguishable from the background. As a side effect, the top electrode additionally attracts DNA from the supply channel. This DNA can either be used as a kind of precharging for the next reactor filling step or be actively pumped out, after the reaction chamber has been emptied (data not shown). If only the electrodes are deactivated, the high concentration of DNA diffuses along the filling channel in both directions, part of which can be seen in figure 4f.

From Simple Single Reactors to Networks and Programmability

The remarkable local control of fluids and concentrations of charged biomolecules shown, and the opportunity to combine this with water droplets in oil or vesicles in water, immediately raises hopes for strong programmability and combining these features in large networks.

Hybrid System Programming and Control

Programming in these systems, in the traditional computer science sense of fixed serial algorithms, is a challenge in several aspects: it has to cope with the continuous versus discrete dichotomy, it involves essentially a stochastic description and it has to cope with reproducibility and error-prone information processing. Even linearization around working points is difficult in many cases. The first attempts to describe optimal control strategies with continuous and discrete systems, termed then as hybrid automata, were presented by Alur et al. [26] in 1993, as a model and specification language for systems consisting of a discrete programme within a continuously changing environment. A major result is that even the very simple and restricted linear hybrid automata verification is already undecidable, although three special cases for which verification is possible are presented. In addition to being a hybrid automata, programmability and control of the chip have to cope with stochastic processes because whether molecules are detected or not depends stochastically on their concentration and on surface properties. This concentration can be either a stochastic continuous variable (10^{14} molecules or more can be seen as continuous reaction fronts that may have arbitrary shapes) or discrete when oil droplets, vesicles or artificial cells are considered. Bujorianu and Lygeros [27] developed a model for general stochastic hybrid systems as a generalization of piecewise-deterministic Markov processes. With the help of dynamic programming, they model general stochastic hybrid systems and solve the derived differential equations. Abate et al. [28] describe stochastic approximations of deterministic hybrid systems, again using Markov processes.

Although hybrid systems research has its origins in classical engineering disciplines, Piazza and Mishra [29] elaborate on questions of stability of hybrid systems in systems biology. Two description languages in the realm of hybrid systems seem to be suitable, CHARON [30] and HYSDEL [31], specifically developed as formal specification languages to programme and control hybrid systems. Using the notation of Alur et al. [26], a typical control element used is shown in figure 5 and will be discussed here briefly.

Low Level-Regulatory Elements

A whole series of low-level control elements have been devised. The state machine diagram of one, the catcher, is shown in figure 5. Two electrodes (actors) and two sensors, suitably placed in the vicinity of the actors, are needed. The sensors can produce two signalling events: a drop in intensity of the fluorescent signal below a certain threshold and surpassing intensity above a second threshold. Both thresholds are programmable, even on-line. If the catcher element has captured a sufficient amount of charged material, it signals a following stage (another low level control element or optionally a more abstract computational entity) that its purpose is finished. The major challenge for incorporating this into control strategies lies in the non-linear, stochastic and error-prone functional relationship between the concentration of material x and the threshold-parameterized output of sensors s0 and s1. In linear hybrid systems, this relationship is defined as being linear and the proofs of optimal control are based on this assumption. It remains to be experimentally proven that, at least for certain operating conditions, these ill-behaved relationships can be approximated by linear functions or that the current research on optimal control including stability analysis can be extended to these types of functions.

Combinatorial Reaction Networks

With these two basic elements, EOF reaction chamber and low-level control elements, the path is open for combining them to create efficient networks. One of the main obstacles in using networks of flow-coupled reaction chambers (see fig. 6 for a design we tested) is to control the flow. Not only did it prove to be difficult to control the flow in the connecting channel of the H structures (because of the imperfections which always exist in the fluid connections and the production process of the imprinted channels in PDMS), it could also be shown that moving DNA packets between the channels do have an effect on the local net flow rates of the fluids in the system (data not shown). With this new strategy, the major flows can now be largely decoupled and independent local cyclic flow systems can be established and controlled. This latter capability will open up a range of possibilities for evolving artificial cells. More practically at this stage, the role of the specially pulse-free (and expensive) external low-flow pumps can be reduced to just feeding the system and letting the electrode systems control the local flow (via EOF).

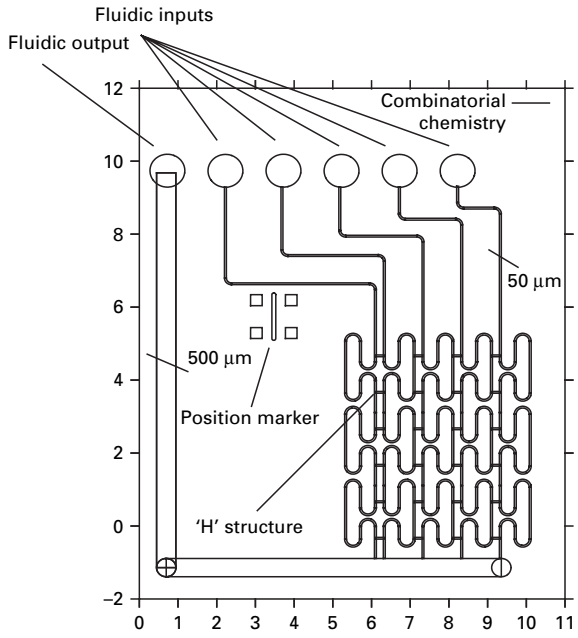


Fig. 6. Test network used to prove the feasibility of combinatorial chemical reactions in microfluidics. The difficult flow control of this system prevented us in the past from doing further research on this topology. However, using the actively controllable reaction chambers shown in figures 3 and 4, in a network presented here (graph of a current fluid design with an n-way fluid crossbar), combinatorial chemistry in micro-flow systems will be feasible. Material from each of the five inputs can be brought into contact with each other. The loops shown are used for delay purposes, to allow a synchronized flow behaviour. The common output channel not only serves to reduce the number of fluid connectors but also as a flow stabilization device (passive regulation).

Discussion and Conclusion

The aspect of functional complementation of artificial cells that has been investigated here is that of electronically regulated and controllable micro-compartments. To support an artificial cell, chemicals have to be introduced to and removed from these compartments and chemical reactions should be sustained for hours. These long-lasting reactions (the metabolic and replication apparatus in true artificial cells is enzyme-free and thus chemical reactions might take a very long time) need to be protected from the outside world and they need to be supplied with the necessary chemicals. Our experimental observation that the transport of charged material worked better for smaller entrance paths has led to a significant enhancement using EOF. Two physical effects help to alleviate the problems of transport. The strong (at least reciprocal cubic) dependency of the hydrodynamic resistance on channel dimension can be used to decouple channels in microfluidic networks ef-

ficiently and drastically reduces fluid noise in the reaction chambers. The second effect is the EOF that utilizes the ubiquitous ionic double layer to produce a net fluid transport under electric fields. Combined, these effects with two-phase systems will further enhance the combinatorial complexity of biochemical reactions which can be regulated in such systems.

The most fascinating potential of this new toolbox is a network of reaction chambers with arbitrary connection topologies and detailed on-line control. Such systems can benefit from and provide new challenges for the decades of experience and knowledge in computer science and dynamical system control. These types of reactor networks can bridge the gap between biochemistry, control theory, computer science and 'hard' artificial life research.

Acknowledgements

This work was conducted as part of the European Union integrated project PACE (EU-IST-FP6-FET-002035). Microfabrication, according to our design, of the EOF reactor by Martina Jünger (RUB) and Thomas Palutke (FhG) is greatly appreciated.

References

1 Sanders GHW, Manz A: Chip-based microsystems for genomic and proteomic analysis. TrAC Trends Anal Chem 2000; 19: 364–378.
2 McCaskill JS, Wagler P: From reconfigurability to evolution in construction systems: spanning the electronic, microfluidic and biomolecular domains. Lect Notes Comp Sci 2000; 1896: 286–299.
3 Ekins R, Chu FW: Microarrays: their origins and applications. Trends Biotechnol 1999; 17: 217–218.
4 Service RF: Will patents flights hold DNA chips hostage. Science 1998; 282: 397.
5 Edman C, Raymond D, Wu D, Tu E, Sosnowski R, Butler W, Nerenberg M, Heller M: Electric field directed nucleic acid hybridization on microchips. Nucleic Acids Res 1997; 25: 4907–4914.
6 Schmidt K, Foerster P, Bochmann A, McCaskill JS: Microsructured flow reactors for two-dimensional experiments; in Microreaction Technology. First International Conference on Microreaction Technology. Berlin, Springer, 1998, pp 233–244.
7 McCaskill JS, Maeke T, Wagler P: US Patent Publication 'Configurable Microreactor Network'. Application No. 09/935,740, 2000.
8 van Noort D, Wagler P, McCaskill JS: Hybrid poly(dimethylsiloxane)-silicon microreactors used for molecular computing. Smart Mater Struct 2002; 11: 756–760.
9 Wagler P, Gohlke M, van Noort D, McCaskill JS: Three-dimensional microfluidic systems for computation with DNA molecules. 12th Micromechanics Europe Workshop MME 2001, Cork, 2001.
10 McCaskill JS: Optically programming DNA computing in microflow reactors. Biosystems 2001; 59: 125–138.
11 Rasmussen S, Chen L, Nilsson M, Abe S: Bridging nonliving and living matter. Artif Life 2003; 9: 269–316.
12 Rasmussen S, Chen L, Deamer D, Krakauer DC, Packard NH, Stadler PF, Bedau MA: Transitions from nonliving to living matter. Science 2004; 303: 963–965.
13 Noireaux V, Libchaber A: A vesicle bioreactor as a step toward an artificial cell assembly. Proc Natl Acad Sci USA 2004; 101: 17669–17674.
14 Tan Y-C, Fisher JS, Lee AI, Cristini V, Lee AP: Design of microfluidic channel geometries for the control of droplet volume, chemical concentration, and sorting. LabChip 2004; 4: 292–298, DOI 10.1039/b403280m.
15 Wagler P, Tangen U, Maeke T, Chemnitz S, Heymann M, Jünger M, Palutke T, McCaskill JS: Dynamical multiphase structures in the microsystem approach to artificial cells. MicroTas 2005, Boston, 2005.
16 Brask A, Goranovic G, Jensen MJ, Bruus H: A novel electroosmotic pump design for nonconducting liquids: theoretical analysis of flow-rate-pressure characteristics and stability. J Micromech Microeng 2005; 15: 883–891.
17 Wagler PF, Tangen U, Maeke T, Mathis HP, McCaskill JS: Microfabrication of a biomodule composed of microfluidics and digitally controlled microelectrodes for processing biomolecules. Smart Mater Struct 2003; 12: 757–762.
18 Duffy DC, Cooper McDonald J, Schueller OJA, White sides GM: Rapid prototyping of microfluidic systems in poly(dimethylsiloxane). Anal Chem 1998; 70: 4974–4984.
19 Ocvirk G, Munroe M, Tang T, Oleschuk R, Westra K, Harrison DJ: Electrokinetic control of fluid flow in native poly(dimethylsiloxane) capillary electrophoresis devices. Electrophoresis 2000; 21: 107–115.
20 Kirby BJ, Hasselbrink EF Jr: Zeta potential of microfluidic substrates. 1. Theory, experimental techniques, and effects on separations. Electrophoresis 2004; 25: 187–202.
21 Liu Y, Fanguy JC, Bledsoe JM, Henry CS: Dynamic coating using polyelectrolyte multilayers for chemical control of electroosmotic flow in capillary electrophoresis. Anal Chem 2000; 72: 5939–5944.
22 Lin C-H, Fu L-M, Lee K-H, Yang R-J, Lee G-B: Novel surface modification methods and surface property analysis for separation of DNA bio-molecules using capillary electrophoresis. MicroTAS, Squaw Valley, 2003, pp 1081–1084.
23 Bianchi F, Wagner F, Hoffmann P, Girault HH: Electroosmotic flow in composite microchannels and implications in microcapillary electrophoresis systems. Anal Chem 2001; 73: 829–836.
24 Mao C, Sun W, Shen Z, Seeman NC: A nanomechanical device based on the B-Z transition of DNA. Nature 1999; 397: 144–146.
25 Tangen U, Maeke T, McCaskill JS: Advanced simulation in the configurable massively parallel hardware: MereGen; in 3rd Caesarium 2000. Berlin, Springer, 2001, pp 107–108.
26 Alur R, Courcoubetis C, Henzinger TA, Ho P-H: Hybrid automata: an algorithmic approach to the specification and verification of hybrid systems. Hybrid Systems, Lecture Notes in Computer Science. Berlin, Springer, 1993, vol 736, pp 209–229.
27 Bujorianu ML, Lygeros J: General stochastic hybrid systems: modelling and optimal control. 43rd IEEE Conference on Decision and Control, Bahamas, 2004.
28 Abate A, Ames AD, Sastry SS: Stochastic approximations of hybrid systems. 24th American Control Conference 2005 (ACC'05), Portland, 2005.
29 Piazza C, Mishra B: Stability of hybrid systems and related questions from systems biology (in honor of Professor Pravin Varaiya on his 65th birthday); in Basar T (ed): Systems and Control: Foundations and Applications. Basel, Birkhäuser, in press.
30 http://www.cis.upenn.edu/group/mobies/charon
31 http://control.ee.ethz.ch/,hybrid/hysdel

Evolutionary self-organization in complex fluids.

John S. McCaskill[1*], Norman H. Packard[2*], Steen Rasmussen[3*], Mark A. Bedau[4*]

[1] Corresponding author: BioMIP, Ruhr-University-Bochum c/o BMZ, Otto-Hahn-Str. 15, D-44227 Dortmund, Germany, Email : john.mccaskill@biomip.rub.de. On leave, Friedrich Schiller University, Jena.
[2] ProtoLife Srl, via della Libertà, 30175 Venice, Italy & Santa Fe Institute, 1399 Hyde Park Rd, Santa Fe NM 87506, USA.
[3] Los Alamos National Laboratory, SOS, EES-6, MS-D462, Los Alamos NM 87545 USA & Santa Fe Institute, 1399 Hyde Park Rd, Santa Fe NM 87506, USA.
[4] ProtoLife Srl, via della Libertà, 30175 Venice, Italy & Reed College, 3203 SE Woodstock Blvd, Portland OR 97202, USA.
[*] European Center for Living Technology, Ca' della Zorza, Dorsoduro 3859, 30125 Venice, Italy.

Abstract

This paper explores the ability of molecular evolution to take control of collective physical phases, making the first decisive step from independent replicators towards cell-like collective structures. We develop a physical model of replicating combinatorial molecules in a ternary fluid of hydrocarbons, amphiphiles, and water. Such systems are being studied experimentally in various labs to approach the synthesis of artificial cells, and they are also relevant to the origin of cellular life. The model represents amphiphiles by spins on a lattice (with Ising coupling in the simplest case), coupled to replicating molecules that may diffuse on the lattice, and react with each other. The presence of the replicating molecules locally modulates the phases of the complex fluid, and the physical replication process and/or mobility of the replicating molecules is influenced by the local amphiphilic configuration through an energetic coupling. Consequently, the replicators can potentially modify their environment to enhance their own replication. Through this coupling, the system can associate hereditary properties, and the potential for autonomous evolution, to self-assembling mesoscale structures in the complex fluid. This opens a route to analyzing the evolution of artificial cells. The models are studied using Monte Carlo simulation, and demonstrate the evolution of phase control. We achieve a unified combinatorial framework for the description of isotropic families of spin-lattice models of complex phases, opening up the physical study of their evolution.

Keywords : *Molecular Evolution, Self-assembly, Ising, Protocell, Amphiphile, Self-organization*

1. Introduction

Physical and chemical models of evolving molecular systems, both theoretical and experimental, have provided important insight into the structure of evolution [1,2]. To date, however, most modeling of evolving chemical systems has been based on chemical reaction kinetics [1], sometimes in conjunction with molecular structure as in RNA landscapes [1,3,4], with spatial effects when considered, studied in terms of reaction-diffusion equations [5,6,7]. The interplay of collective self-assembling mesoscale structures [8,9,10] with the evolution of molecules has not yet been investigated in a coherent physical model. The current work aims to bridge this gap, and thereby to open up the joint investigation of dynamical self-organization with a single model exhibiting coupled self-assembly and evolution. Models of amphiphilic self-assembly have been much investigated [11,12], not least because of their relevance to cellular membranes, but although the relations between molecular properties (such as head group size and chain length) and the equilibrium phase diagram are now quite well understood, the evolution of molecules utilizing collective phases based on this relation is not. The evolution of genetic molecules with surfactant properties was proposed to be important for the origin of life[13]. The physical connection between evolution and self-assembly in complex fluids is the topic here.

Rather than starting with a simulation of a physical abstraction of a particular protocell[14,15], we seek to explore the general class of evolvable multiphase systems in a physically grounded model. This work extends insights from evolving spatial reaction diffusion systems [7,16,17] to such multi-phase fluids, where mesoscale structures create mobility barriers and local thermodynamic reaction conditions for replicators. In terms of evolution theory the question is how different replicators can cooperate to induce appropriate collective phases for their proliferation. The effects of mesoscale structures on molecular evolution is bound to be important, because it is now well accepted that spatial localization of chemicals, even in simple reaction diffusion systems, enables complexity in evolving systems[16,18]. In particular, the Gray-Scott mechanism of resource limited cooperative replication was shown[19] to support growing and dividing spot patterns, reminiscent of cells, and these structures actually in turn promote the evolutionary stability of the cooperative replication. In any case, the importance of cell-like mesoscale structures for biological evolution need hardly be stressed, so our goal is to reduce the model complexity to the point where the key processes and phenomena can be illuminated by simple, but still grounded physico-chemical models and simulation.

Our model is intended to remain as simple as possible while retaining the essential features of an amphiphilic system under the control of a combinatorial family of replicating molecules. To this end, for the amphiphilic system, we base our model on the simple spin-lattice models introduced by Ising [20] and Widom [21,22], which we shall however generalize to embed them in the context of a general evolvable family of isotropic models with local interactions. The Ising model will serve as a simplest reference system for highlighting the basic physics entailed in coupling collective physical phases with evolving molecules, towards life-like systems and demonstrating that important issues of collective evolutionary self-organization can be addressed in the simplest models of phase transitions. The Widom case establishes the applicability of these considerations to amphiphile systems relevant to emerging life. We will study our model in two and three dimensions: in 2D for visual simplicity and computational speed and 3D for more physical realism and topological variety.

In addition to the chemically simple amphiphile-hydrocarbon-water system, we assume that a combinatorially complex class of molecules, capable in principle of self-replication, may also be present. These combinatorial molecules are assumed to influence the free energies of local amphiphile configurations, and consistent with this influence, their diffusion on the lattice is kinetically biased by these induced free energy changes. If we include some specific influences of the local phase on the replication or degradation processes for these replicators, we arrive at a rich thermodynamically consistent model of amphiphile-phase dependent and phase-influencing replication that may serve as an elementary physical model for studying the evolutionary transition from chemical to cellular systems. Thermodynamic consistency is important if we are to ensure a physically sound model of information flow in the interplay between self-assembly and catalytic self-organization.

2. The Evolving Self-Assembling Phases (ESP) model

In this section we first introduce the basic homogeneous, spin-lattice thermodynamic models and their extensions to a family of more general isotropic hamiltonians. We then discuss our choices in formulating kinetics (non-equilibrium treatment) for these models, before describing our main innovation of heterogeneously modulated spin lattice systems. The local modulation of thermodynamic parameters is

attributed to the presence of a family of combinatorial molecules, which are introduced as initially static, then as mobile diffusing and finally as reacting and self-replicating entities.

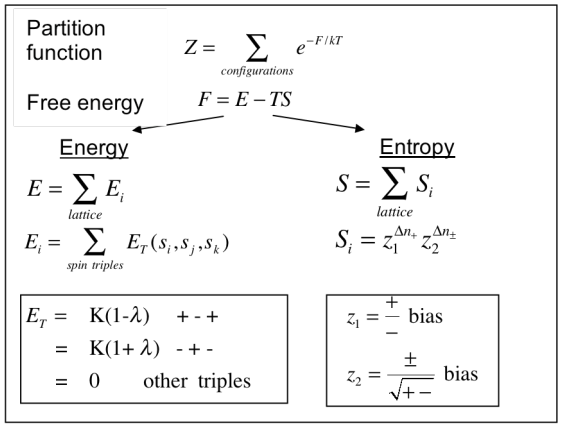
The basic Widom spin-lattice ternary model

The underlying amphiphile model we have chosen is motivated by the success of Widom's spin lattice model and its relatives in modeling the phase diagram of ternary fluids. The Widom model[21] reduces the complexity of ternary fluids comprised of hydrocarbon, water and amphiphiles to a two state model, bringing it remarkably close to the simple Ising model, which can also be applied to fluid mixtures.

The Widom model is based on a regular square or cubic lattice of sites with two states (+, -), which represent a collection of either hydrophobic or hydrophilic ends of molecules meeting at the site. In this model, all species are modeled as dimers, straddling two sites of the lattice, with hydrocarbon (oil) molecules (-, -), water (+, +) and amphiphiles (+, -) or (-, +). Because of their prohibitively large energy, configurations where a mixture of hydrophobic and hydrophilic ends meet are excluded. This is reasonable for systems containing amphiphiles, where oil-water interfaces will always be mediated by amphiphiles. Although the model cannot well describe isolated amphiphiles in water and hence the critical micelle concentration, it has proved effective in reproducing many features of the phase diagram of oil-water-amphiphile emulsions. The model is grand canonical, such that both material and energy can be freely exchanged with the environment. In particular, activity coefficients are used to specify the concentration biases of water to oil and amphiphiles to water and oil. For example, changing a local "spin" state may induce a change in the number of amphiphiles in the system.

< Fig. 1 Here >

The equilibrium thermodynamics can be derived from the partition function, which is defined in Box 1. The Widom model involves a mean energy difference K when the ends of two amphiphiles meet at a lattice site, rather than two ends of oil or water dimers, reflecting differences in the interactions of amphiphiles compared with water or oil. Furthermore, the hydrophobic ends of amphiphiles can have a different energy $K(1+\lambda)$ from the hydrophilic ends $K(1-\lambda)$. The relative activity coefficients, z_1 and z_2, describe respectively the entropic cost in the grand canonical ensemble of number fluctuations away from equal occupation by water vs. oil, Δn_+, and amphiphile vs. homophile, $\Delta n_\pm$.



Box 1

Basic thermodynamics of the Widom model for the spin configurations explained in fig. 1. The basic model has four parameters K, λ, z_1, and z_2. Note that for the homogeneous model it is sufficient to either vary K or the temperature T to obtain the full range of behaviors of the model.

A general evolvable local isotropic Hamiltonian

The Widom model is a particular instance of a local energy model for an amphiphilic system, assigning energies to amphiphilic spin configurations consistent with translational, rotational and reflection symmetries of space. In fact, examining the most general model with nearest neighbor interactions, with a view to later permitting the evolution of arbitrary local amphiphilic interactions, we find a reduced number of distinct local spin configurations, subject to spatial symmetries depending on the lattice. Although the situation is simpler on a square 4-neighbor than a hexagonal 6-neighbor lattice in 2D, we have used the hexagonal lattice for some of our simulations because of known artifacts on a square lattice. In 3D, the greater number of neighbors appears to allow enough configurational flexibility even for a cubic lattice.

Extensions to allow longer chain amphiphiles, as in the Larson model [23], which allow further mesoscale structures such as vesicles, as well as for an explicit presentation of amphiphiles, see e.g. [24], are postponed to subsequent investigations. Extensions to include higher order spin terms (e.g. four spin) with next to nearest neighbors, allowing the alignment costs for amphiphiles to be considered, while of interest in capturing certain properties of the phase diagram, are clearly in a later round of investigation for evolutionary models. Thus while the Widom model is known to have certain limitations in describing some

aspects of the ternary phase diagram, essentially these simple spin-lattice models are known to explain the main features of the progression of phases in micellar systems with oil, water and amphiphiles[21].

In a general isotropic (nearest neighbor) lattice model of this kind, the energies are determined by the interactions between adjacent spins and spin triples, independent of their detailed relative orientation. Allowing spatially symmetric additive contributions to the energy, up to spin triples, with no distinction between different geometric arrangements of triples, the local energy depends on six parameters:

$$E[\underset{\sim}{x}] = a_1(\eta_1 + \eta_2\sigma[\underset{\sim}{x}]) + a_2(\eta_3 + \eta_4\sigma[\underset{\sim}{x}])\sigma_{sum}[\underset{\sim}{x}] + a_3(\eta_5 + \eta_6\sigma[\underset{\sim}{x}])\sigma_{skew}[\underset{\sim}{x}] \quad (1)$$

where $\sigma[\underset{\sim}{x}]$ is the central spin at the lattice site $\underset{\sim}{x} = \{x, y, z\}$ in 3D or $\underset{\sim}{x} = \{x, y\}$ in 2D and

$$\sigma_{sum}[\underset{\sim}{x}] = \sum_{neighbours\ i} \sigma[\underset{\sim}{x}_i], \quad \sigma_{skew}[\underset{\sim}{x}] = \sum_{\text{neighbours } i>j} \sigma[\underset{\sim}{x}_i]\,\sigma[\underset{\sim}{x}_j] \quad (2)$$

The parameters a_k are positive integer constants depending on the lattice: for the square lattice (6,3,1) and for hexagonal (2D) and cubic (3D) lattices (15,5,1).

For the Ising model, the only non-zero values of η are η_2 and η_4, equal to the Ising field strength H and the spin coupling constant J. In the case of the original Widom model, the parameters η_l above are

$$(\eta_1,\ \eta_2,\ \eta_3,\ \eta_4,\ \eta_5,\ \eta_6) = (K/4) \times (1,\ \lambda,\ -\lambda,\ -1,\ 1,\ \lambda). \quad (3)$$

To illustrate the extended formula, the Widom parameters above can be used to calculate the local energy contribution of the central site in the configuration given in Figure 1 as follows (hexagonal 2D lattice):

$$E = 15\frac{K}{4}(1+\lambda(-1)) + 5\frac{K}{4}(-\lambda - 1(-1))(0) + 1\frac{K}{4}(1+\lambda(-1))(-3) = 3K(1-\lambda). \quad (4)$$

The energy contribution is just that for 3 triples with hydrophobic ends of amphiphiles meeting at site, as expected for the Widom parameter special case.

If we completely differentiate configurations (equating the energies of states identical under rotation and reflection) then we arrive at a somewhat larger number of different energies ς_i for the general nearest neighbor Hamiltonian (not restricted to lower order spin terms), depending on the lattice: square (12), hexagonal (28) and cubic (20). The number of parameters is still significantly reduced compared with the number of local configurations (32,128,128) by the physical isotropy requirement. In addition to the interaction energy parameters, the lattice models contain parameters defining the grand canonical ensemble (either activity coefficients or chemical potentials): z1 for Ising and {z1, z2} for Widom. While these have also been subject to genetic modulation, we defer this to later investigations.

Kinetics of the multiphase system

Although the models are amenable to extensive theoretical analysis in equilibrium, using for example mean field, self-consistent field and transfer matrix techniques, and this allows a characterization of the phase diagram for the homogeneous case, our primary interest in this paper is in the kinetic perturbation of these equilibrium phases by the combinatorial molecules and so we need to incorporate kinetics.

Arguably the simplest kinetics consistent with the above partition function is a Monte Carlo (MC) model[25] with single spin flips resulting in a free energy change ΔF_1 being accepted with probability $P_1 = \mathrm{Min}(1, e^{-\Delta F_1/kT})$. Such a single spin flip dynamics is equivalent to collectively exchanging all halves of dimeric molecules at a lattice site. In particular, this means for example that spin flips can cause amphiphiles and then water to appear in the middle of an oil droplet. This makes some sense for an open system in 2D, if we imagine the exchange of molecules with an off-surface reservoir for example, but is less easy to justify in 3D, unless we view the system to be in continual turnover in connection with reservoirs of dimeric molecules of the various types. Restricting the dynamics to interchange of neighboring pairs of spins [26] would conserve the number of hydrophilic and hydrophobic molecule ends, but not the number of amphiphiles. The kinetics with such a semi-conservative dynamics tends to be much less efficient in equilibrating the system, and so we have opted for the more efficient dynamics in this work, while aware of its limitations.

For the single spin flip MC dynamics, the resulting energy difference must be computed for the central site and it's nearest neighbors. The change in the number of oil, water and amphiphilic dimers induced by the spin change must be computed only at the central site, to determine the entropic cost of the transition associated with the activity coefficients. The resulting free energy differences (see Box 1) can be computed efficiently. The traditional spin formulation can be replaced by a neighbor spin table lookup, enabling the physical transition probabilities with detailed balance to be considered in the broader context of stochastic

cellular automata update rules. In particular, the traditional additive energy contributions from combinations of small numbers of spins can be extended to nonlinear combinations of spins of higher order than the nonlinear triplet terms.

Combinatorial energetic variation coupled to the multiphase system

Because of the ephemeral nature of amphiphiles in the Widom model, it is not productive to try to attach combinatorial information about interaction properties to amphiphile configurations. The number fluctuations would result in a rapid loss of this information, or require artificial reservoirs to keep track of the dynamics of hidden molecules. Instead, we chose to consider separate combinatorial molecules that modulate the hydrophilic interaction properties (e.g. K and λ, or z1 and z2) at a lattice site, thereby influencing the amphiphile system, but not subject to rapid number fluctuations. This has the additional advantage of separating the time-scale for combinatorial change from the relaxation time of the amphiphile system. In the discussion below, we will also refer to these combinatorial molecules as *templates* or *replicators*. Portions of these molecules encoding parameters that specify the modulation of the amphiphilic interactions (e.g. K and λ) at a given lattice site are referred to as genes.

<Fig. 2 Here >

A static population of such combinatorial molecules corresponds to a heterogeneous spin-lattice amphiphile system model, with fixed but spatially varying energy (and/or entropy) parameters at each site. In two dimensions, this could correspond to a surface local modulation of hydrophilicity by a solid support. The result is a generalization of the spin-glass random energy model for disordered solids[27], but is not necessarily of special interest in the context of fluid amphiphilic systems. Below we shall therefore allow the combinatorial molecules to diffuse. Regular gradients of the genetic parameters can be employed in both two and three dimensions to scan a whole range of parameter conditions for interesting local phases. Such fixed multi-dimensional gradient scans have already been found useful for reaction diffusion systems such as the Scott Gray system[19], and were found useful here in tuning the starting parameters of our model. We have set up our model to allow genetic parameters to potentially influence z_1 and/or z_2 and one of J for the Ising model, K and/or λ, in the Widom case, a subset of the six extended Widom parameters η_l in the isotropic triple spin extension, or the 12, 28 or 20 isotropic site energy parameters ς_i (for square, hexagonal or cubic lattices, respectively) in the full nearest neighbor spin model. We typically consider a linear range of variation between maximum and minimum values. This is also the range of modulation we shall investigate when allowing template mobility and evolution as described below.

Figure 2 shows two typical gradient images of a simulation of the amphiphile spin-lattice system obtained for the (non-extended) Widom model itself. If we let the simulations run long enough, and make the system large enough, the gradient portrait is equivalent to an equilibrium phase diagram. Note the transitions between homogeneous, micellar, lamellar and liquid crystalline phases which are observable in the model. The micellar interpretation is problematic in the Widom model (they are actually small emulsion droplets), which also cannot exhibit the CMC transition, because it does not allow for isolated amphiphiles. Note also that the gradients must be very weak to be able to infer accurately from the local phase at a location in the gradient field, what the phase of the homogeneous amphiphile system with these local parameters would be. This is especially true for gradients in the activity coefficients ln z_1 and ln z_2 since spin configurations can appear at one location and migrate (by single spin flips) to another. Nonetheless, the gradient simulations not only give a rapid indication of the equilibrium phase behavior of the system, but also seem to capture some transient phenomena resulting from heterogeneity, which is likely to be relevant in the context of evolution.

Mobile combinatorial molecules

If we allow the combinatorial molecules to diffuse, then we have a novel mobile-fluid phase model. However, it appears at first sight unavoidable that allowing such diffusion will break the thermodynamic consistency of the amphiphile model. The problem is that diffusion of such molecules induces free energy changes in the amphiphile system through their modulation of its local energetics, and these changes need to be reflected in biasing movement. These free energy changes must be taken into account for diffusive interchange of a pair of neighboring combinatorial molecules, which we do using a Monte Carlo weighting as above. This means that for a fixed amphiphile (spin) configuration, the free energy difference, ΔF_2, on interchanging the combinatorial molecules (thermodynamic parameters) between two sites, which may

have either sign, is calculated (including the energetic effects induced at neighboring sites), and the interchange accepted with probability $P_2 = \mathrm{Min}(1, e^{-\Delta F_2 / kT})$.

The advantage is now that we have a very natural and thermodynamically consistent feedback of amphiphile system information controlling the microscopic mobility properties of a combinatorial family of molecules. This is in fact what is required if we are to move beyond the simple pattern formation properties of reaction diffusion systems as proposed by Turing[28] to include the possibility of evolution. In cell biology this model might also be used to represent the biased diffusion of protein molecules embedded in a lipid bilayer membrane.

Template chemistry of combinatorial molecules

Chemical reactions among the combinatorial molecules allow a dynamical self-organization of the phase behavior of the amphiphile system to take place, and in turn the amphiphile system can modulate the chemical reactions, both through influencing the reaction rates directly via the local phase or indirectly through the influence on template mobility and effective mutual affinities via the induced amphiphile energetics (as above). We shall want to include chemical reactions that allow the proliferation and destruction of combinatorial molecules, which we shall then call replicators. Chemically no complex biochemical apparatus is required for replication, as shown by the known experimental examples of enzyme-free template chemistry such as [29], although to date only limited combinatorial complexity has been achieved without enzymatic mediation. Non-explicit cooperative effects can occur even in the simple replication mechanism mediated by the amphiphile system.

Simple error-prone replication ($X \to 2X$) of a combinatorial family of molecules X corresponds to the well-studied [1] physico-chemical model of the quasispecies introduced by Eigen[30]. The cooperative replication mechanism ($2X \to 3X$) is a standard model for the problem of functional exploitation[16] and its spatial resolution, which is germane to the origin of life and generates proliferating cell-like reaction-diffusion patterns (self-replicating spots) [19], which have been shown to allow evolutionary stabilization of the cooperative functionality with a much simpler mechanism than that of higher order hypercycles[31] via spirals[5]. In the current investigation it is of primary interest to investigate the indirect cooperative effect via the amphiphile phase system without such explicit higher order catalysis.

Note that detailed mechanisms of templating, e.g. via polymerization from monomers or ligation of fragments, are not considered in this work. In particular, the common[32] chemical replicator model ($X_a + X_b + X \rightleftharpoons X_a X_b X \to X_2 \rightleftharpoons 2X$) with parabolic growth and coexistence of multiple templates is not investigated here, since we are interested in the evolution of combinatorially more diverse replicator functionality. The phenomenon of coexisting templates itself is also not so interesting, despite much recent attention, being a property of the simpler self-inhibition by double strand formation $B + X \to X_2 \rightleftharpoons 2X$ (with or without the resource B) as shown in[33], and studied for ligation[34] abstracting Blum's model[35].

To connect replication with the amphiphile system of our model three things are necessary. Firstly, we need to translate the mechanisms into nearest neighbor configurational changes on the lattice. We consider all replicators X as existing on the sites of the amphiphile lattice, occupying a lattice site with at most single occupancy, with default amphiphile energetics being employed in their absence. Replicators, X modulate the energetic properties of the amphiphile lattice at the site they occupy. When new combinatorial species are created, they are placed at a neighboring lattice site, overwriting the original site occupant. Secondly we need to preserve thermodynamic consistency of the amphiphile system, by testing reaction outcomes thermodynamically. Placing at a neighboring site a replicated combinatorial molecule, affecting some subset of local amphiphile thermodynamic parameters (e.g. ζ's and z's), induces a free energy change ΔF_3 on the amphiphile system, by the same local thermodynamic contributions described above. As above, the reaction can be accepted or rejected via the Monte Carlo probability $P_3 = \mathrm{Min}(1, e^{-\Delta F_3 / kT})$. Thirdly, direct influences of the local phase on the reaction can be included optionally: the reaction taking place only if the local site is hydrophobic or hydrophilic, or only at the interface between like or different phases.

To prevent complete saturation, the chemical destruction of combinatorial molecules is also considered. In this work, all modulators are assumed to be destroyed at a constant rate, but at a rate that may depend on the amphiphile phase. In addition to the possibilities of specific destruction rates in either the hydrophobic (-) or hydrophilic (+) phases, we also may consider the possibility of specific destruction either at an interface (a site with an oppositely signed neighbor) or in bulk (the reverse). Although we have begun exploration of a broader range of reaction behaviors in our model, in this paper we concentrate on the simplest scenario in which not only destruction but also replication rates are not directly sequence

independent. This highlights the role of the complex fluid coupling in determining the evolutionary outcome. Antagonistic scenarios where cooperation between replicators has to overcome differential non-cooperating proliferation have been shown elsewhere[17] to be kept in check by spatial correlations in a manner similar to the neutral case adopted here for simplicity. We have also chosen the simplest independent replication mechanism without specific resources, and considered just two simple scenarios where sequence independent replication is allowed either in the hydrocarbon phase or at the interface.

Combinatorial encoding of amphiphile modulation, reaction rates and mutation

In order to provide a general modeling framework we divide the genetic effects into two classes, those affecting the amphiphile thermodynamics and those affecting reaction rates. The first class encodes both the local amphiphile energetic parameters and the two chemical potentials. The second class specifies reaction probabilities, typically just one such as for replication, and this is not investigated in this work. The thermodynamic values are encoded in binary sequences, abstracting more general combinatorial structure, and interpolate linearly between maximum and minimum values. The ranges for the signed parameters ς (or z) are given as parameters to the simulation in the form $\varsigma = \varsigma_0 \pm s\Delta\varsigma$, where s is a finite precision signed binary fraction encoded by the sequence, between -1 and 1. Sixteen bits were used per parameter for z_1 and z_2, and for K and λ in the Widom case. For the extended parameters ς_i, bit lengths between 2 and 5 were used, depending on the lattice (5 for square 2D lattice, 2 for hexagonal lattice).

Mutation is performed as single bit substitutions with a constant uniform probability. We allow multiple mutations with exponentially decreasing probability, by repeating the mutation decision for a sequence until it fails. Note that while the extended parameters ς_i can be restricted to values equivalent to the original Widom case, (yielding parameters K and λ), their mutation must then also be restricted in order to study evolution within the restricted Widom model.

Simulations and software

The computer code was written jointly in Mathematica™ 5.2 (Wolfram Research) and the GNU GCC C/C++ compiler, using the Mathlink connection. This allowed algebraic computations of parameters to be interfaced with optimized runtime code, and provided powerful high level graphics and GUI handling capabilities. The simulations were performed on G4 and G5 Macintosh processors under OS X 10.4, with typical equilibrating CPU runtimes on the order of seconds to minutes, and evolution runtimes on the order of minutes to an hour. The graphical display language SDL was used for intermediate graphical display.

Discussion : Coupling of system thermodynamics and kinetics

The presence of a genetic molecule locally modifies the free energy of the complex fluid due to the hydrophobic and hydrophilic (or amphiphilic) interactions and Newton's 3rd Law dictates that this interaction has to be symmetric so that the ternary system locally influences the genetic molecule in the same manner. If, for instance, a template is amphiphilic, it will tend to locate itself at a nearby water-oil interface if possible, and have a lower probability of leaving this state. Consequently, the combinatorial molecule diffusion is phase dependent, since for example an amphiphilic combinatorial molecule is much more likely to diffuse along an oil-water interface and much less likely to diffuse into the oil or into the water phases. Conversely, the presence of the amphiphilic combinatorial molecule will influence the local energetics of the substrate to produce more local interface if possible. The gene-phase interaction strength can be adjusted in the model using the parameters $\Delta\varsigma_i$, since these specify the maximum magnitude of sequence dependent physico-chemical properties of the combinatorial molecule on the local entropic or energetics of the complex fluid.

Further, it is assumed that the genetic molecules' ability to replicate and/or survive depend on their environment. E.g. a hydrophilic combinatorial molecule will replicate with a much higher probability in an aqueous local environment. These catalytic genetic properties are reduced to a case-by-case study in our model: the combinatorial molecules can replicate (either at the same rate or at a rate determined by their sequence) only in a predetermined local phase: either oil, water, interface, bulk, or ubiquitously. Mixed scenarios are not considered at this stage. Attention here is devoted to the neutral case, where direct sequence dependent effects on replication rate are absent: a combinatorial molecule can only influence its survival indirectly via interactions with the complex fluid phases.

Since the underlying multiphase system is represented in a grand canonical ensemble there is no synthetic cost associated with mass transport of water, oil, or amphiphilic molecules into and out of the multiphase system. This means that the local energetic influence that a combinatorial molecule imposes on the complex fluid can in a metabolically neutral manner drive a local mass transport of water, oil, and amphiphiles. This is equivalent to assuming that the complex fluid components are freely available or readily synthesized and do not represent combinatorially complex entities. Each combinatorial molecule thereby slightly modifies the local fluid composition and thus the local phase. This resulting phase will in turn influence the combinatorial molecule mobility, the location, as well as the replication rate, see fig. 3.
The combinatorial molecules are not represented in the grand canonical ensemble, because the heredity of their information is a definitive characteristic limiting evolutionary survival. Since there is free energy exchange between the complex fluid and the combinatorial molecules, the combinatorial molecules are represented in the canonical ensemble. Since our simplified model does not explicitly include any metabolic processes and only the combinatorial molecules are conservatively synthesized by the system, genetic replication and loss defines the non-equilibrium energetic pumping of the overall system. The overall system consists of the template molecules coupled with the ternary fluid system. Each replication event is associated with a free energy cost and this (arbitrary, but fixed) free energy is supplied to the system from the outside. We can safely assume that the free energy associated with this pumping (per molecule) is much greater than the free energy associated with the relaxation and self-assembly processes occurring in the multiphase system (per molecule), as the former is associated with formation of covalent bonds while the latter is associated with the hydrophobic effect.
Due to these gene-ternary fluid interactions, evolutionary adaptation happens in two fundamentally different ways in this system: (i) In the manner that is traditionally considered where the generated variation in the template population, due to random mutations, enables a selection for templates with the fastest replication rate for a give environment, which is normally called *adaptation to the local environment*. (ii) The combinatorial molecules can also be selected for as they *adapt the local environment* to enhance their own replication rate. This is possible because the presence of the templates modify the local thermodynamic conditions of the ternary system, which in turn modifies the thermodynamic (and kinetic) conditions for these combinatorial molecules. Thus, particular templates can be selected for, which modify the local environment, which in turn enhance their replication rate.
This adaptation of the local environment is a collective (cooperative) effect (of multiple genetic molecules) because of the way local interactions collectively induce phases in the complex fluid. Such population dependent interactions and environmental influences are known potentially to complicate Darwinian evolution, producing outcomes other than simple adaptive optimization (including limit cycle and more complex oscillatory or chaotic evolutionary dynamics). Spatial effects then clearly play an important role, in the feedback between genetic influence on the complex fluid and its influence on genetic survival.

< Fig. 3 Here >

3. The combinatorial self-assembly dynamics without reactions
Before considering any reactions or evolution of the combinatorial molecules, it is important to gain some understanding of the nature of this coupled dynamical system: how it affects the amphiphile phase diagram and how it impacts on the free diffusion of modulators. This will be the subject of this section.
Although we shall demonstrate this two-way coupling using generalizations of the Widom model, and its essential precursor, the random coupling Ising model, we expect the type of phenomena we demonstrate here to be a general feature of combinatorially coupled complex fluid models.

< Fig. 4 Here >

Note that the influence of random static, non-reacting combinatorial molecules is related to the random energy models as introduced in the study of spin glasses and other disordered solids[27]. The simplest case is just an Ising model with a random site energy (or single spin H parameter) or random coupling terms J. In this static case, it is known that the disorder itself can induce localization thresholds in the mobility of particles subject to this Hamiltonian, but this is unlikely to be the case if the random energies are themselves mobile. In the case of random H, the perturbations on the Ising model behavior appear local. In the case of random J, the critical temperature itself can be shifted radically by random perturbations as shown in the gradient plot in Figure 4. In order to demonstrate the main effect we chose reference systems

with zero interaction energies. In the lower plots, static inhomogeneous perturbations of the Ising model, as applicable to a binary fluid, are shown. Note that in the absence of the fixed combinatorial molecules there is no ordered phase at any temperature, since J is set to 0. The correlated patches in the ordered regime are at fixed locations in the random J model. The figures were created with Monte Carlo simulations starting with a random spin configuration.

A similar result is achieved for the Widom model with fixed combinatorial molecules perturbing $K^+ = K(1+\lambda)$ and $K^- = K(1-\lambda)$, as shown in the upper right of Figure 4. Note that it is more natural to treat these two quantities as the independent random variables subject to genetic control, than K and λ directly, because they represent respectively the separate energies of the juxtaposition of hydrophilic or hydrophobic ends of two amphiphiles ends, compared with water or hydrocarbon. Although these plots are noisy, being subject to the uncorrelated perturbations of the combinatorial molecules, more correlated perturbations induced by the combinatorial molecules can have a stronger influence.

< Fig. 5 Here >

If we now allow the combinatorial molecules to diffuse, as befits disordered fluids, in contrast with solids, then a number of new phenomena emerge. Recall that the random walks must be biased by the energetic changes induced in the binary (Ising) or ternary (Widom) fluid for thermodynamic consistency, as discussed in section 2. Consequently, the mobile combinatorial molecules can both influence the phase of the complex fluid in which they float, and their spatial arrangement and mobilities can be influenced by the induced phases. In Figure 5, we show this coupled phase behavior for the Ising model coupled to mobile elements which randomly perturb the coupling constant J from zero. This describes the way in which a random population of combinatorial molecules with varying degrees of amphiphilicity interacts with a homogeneously mixed two component fluid to induce mesoscale structuring of the fluid and combinatorial sorting and clustering of the combinatorial amphiphiles. Note that the combinatorial molecules themselves may experience effective attractive or repulsive interactions allowing aggregation for example, induced by their impact on amphiphile energetics. We note that such examples of noise induced order may assist the onset of cooperative action by populations of genetic elements, through correlating collocation with function autonomously.

< Fig. 6 Here >

A similar non-random clustering of the diffusing combinatorial molecules, interacting with a Widom ternary emulsion system of hydrocarbon, water and surfactant, occurs as shown in Figure 6. So, in addition to the vertical gradient, we also show a horizontal gradient in the activity coefficient of amphiphile (z2). Near z2=0, at the left and right extremities of the plot, the higher temperature phase is the mixed system, whereas at lower amphiphile activities, in the center, the higher temperature phase is the separated fluid. The spatial resolved inverse mobility plot in Figure 6b confirms the structured feedback on the combinatorial molecules through their perturbation of the complex fluid. The values shown in Figure 6b are calculated as the free energy cost in units of kT of an interchange of two neighboring amphiphiles in the vertical direction. The plot looks similar for interchanges in the horizontal direction. The mobilities of the combinatorial molecules become strongly correlated in spatial clusters of molecules having similar interaction parameters below the transition to the higher temperature phases.

< Fig. 7 Here >

In Figure 7 we show the corresponding free energies, which drive the equilibration process for the Widom model case shown in Figure 6. Note that equilibration is essentially complete after 1000 Monte Carlo updates per site. It is worthwhile to observe that the combinatorial molecules relax to lower the free energy of the system, at the expense of the configurational free energy associated with the ternary fluid, consistent with the equilibrium state containing more structure than in the absence of the combinatorial molecules.

In summary, in this section we have shown that mobile combinatorial molecules can both modulate the phase of complex fluids and themselves be ordered by this interaction. It is significant that random perturbations by combinatorial molecules can induce phase transitions to more ordered phases of complex fluids (as in noise induced order). Note that although static random energy and random coupling models have been studied in solid state areas such as spin glasses, the equivalent for a multiphase liquid system

with mobile elements does not appear to have received attention until now. The ability of molecular interactions with a complex fluid to provide structured pattern formation in a system of combinatorial molecules introduces a rich range of novel pattern forming mechanisms to act as a substrate for evolution. This will be explored in the next section.

4. The evolving replicator-amphiphile system

In this section we shall explore the coupled evolution and self-assembly which our model captures. This is accomplished by allowing reactions to take place among the combinatorial molecules, in particular, replication, which enables inheritance and subsequently evolution. The simplest case is that of the Ising model as applicable to a simple binary fluid, as discussed in Section 3, which we shall examine first. This exemplifies the character of evolution coupled to phase control in a thermodynamically consistent way. The second case is the Widom model, and because it incorporates a combinatorial treatment of amphiphiles, this of course is more relevant to the self-organization of protocells. Both models are special cases of the general isotropic local Hamiltonian introduced in Section 2, which will in the future be investigated for more general evolution.

Consider firstly the evolution of a population of combinatorial molecules which can only replicate in an organic phase, but which can influence the coupling constant J, determining the interaction energy between the organic molecules and water, in a sequence dependent fashion. Since non-zero coupling constants can support collective phases, it is conceivable that templates with sequences for appropriate values of J, can cooperate to produce a local oily phase and thereby increase the replication probability, and survival of these templates. As in Section 3, we setup a temperature gradient environment (from 0.25 to 2.25), with the environmental coupling coefficient J_0=0, so that the initial fluid is in a well-mixed state containing equal amounts of water and organic material. The range of combinatorial variation is set to ± 1. There is no combinatorial variation in replication probability or loss rate, which are set to 0.5 and 0.1 per MC update (unit time). We use simple replication and the binary landscape described in Section 2 and a single mutation rate of 0.01 (per replication event, with multiple mutation events per replication allowed). The results are shown in Figures 8 and 9. Note that at low temperatures, the steady state population supports some “parasitic” combinatorial molecules, which do not contribute in full to maintaining the oily phase (Figure 8e).

< Fig. 8 Here > < Fig. 9 Here >

The evolution of combinatorial molecules capable of acting as or producing amphiphilic molecules will be the subject of the remainder of this section. This is of some significance to the evolution of cellular life, as suggested by the experiments on self-reproducing surfactant systems[36]. Already in the Ising model simplification, some useful insights can be gained, so we present an evolutionary simulation before proceeding to our main result on the Widom model. If replication of template molecules can only occur at the interface between hydrophobic and hydrophilic phases, an assumption relevant to the Los Alamos protocell model [14], then the situation changes dramatically from the replication in oil scenario shown above. We commence with a finite environmental value to the coupling constant J=0.5, so the system is phase separated at nearly all temperatures in the gradient simulation (T from 0.5 to 1), see Figure 11a. We allow a population of combinatorial molecules to influence the coupling constant as in the previous example. Although we could have attained more rapid evolution with the binary sequence landscape used in the previous case, we have used a unary encoding of the coupling constant here (30 bits), which serves to illustrate a landscape where significant functionality is rare and where many mutations from random sequences are required to reach this.

< Fig. 10 Here >

The evolution shows increasing adaptations of the combinatorial molecules to producing interface via negative J values and thereby an emulsion of the oil-water system. Initially, the population is localized near the macroscopic interface, with minor departures determined by the relative magnitudes of diffusion and loss probabilities. Interface is more easily come by at high temperatures, so the time course shows a successive occupation towards the low temperature center. The steady state evolved population occasionally completely fills the space, but predominantly a small phase separated region at low temperature remains (Fig. 11c). On average, 5 adaptive mutations are required to achieve the coupling

values of the evolved population, with more negative values being adapted in the central low temperature region (Fig. 11 e). Compare this with the neutral case, where local spatial correlations induced by the replication-diffusion system are visible (Fig. 11f). The free energy traces Fig. 10 d-f, shows that the evolution drives the system steadily, but non-monotonically, to a higher free energy state with about half the free energy in the genetic modulation.

< Fig. 11 Here >

The final example discussed in this section demonstrates the evolutionary adaptation of replicators' modulator genes to enhance their proliferation via genetic control of self-assembled amphiphile structures. The example is not just selection from among the variants present in an initial population, but a sequence of at least two separate adaptive events. The first occurs some time after the simulation starts, when a colony of replicators first successfully adapts to the watery phase. This adaptation changes the water phase by creating lots of micelles, thus maximizing interface (thus, maximizing reproduction). The second, later adaptation arises independently in a number of new colonies that learn to survive in the oily phase by creating reverse micelles, thus maximizing interface. The timing of these adaptations indicates that they were not present in the initial population (or were present but did not survive).

< Fig. 12 Here >

The example uses an environment created by a gradient of z_1 ranging from –2.7 to 2.7, which creates two bulk phases (water and oil) with a complex interface between them. The replicators evolve different adaptations to survive in these niches, as well as at the interface. Replicators are chosen to replicate only at the interface. Other thermodynamic parameters are $kT = 0.2, z_2 = 1.3, K = 0.25, \lambda = 0.01$. Figure 12 illustrates the approach to equilibrium behavior of the amphiphile system with no replicators present in 12(a, b). In figure 12(b), we see the watery phase on the right (white), the oily phase on the left (black), and complex phase in the middle with a few micelles (black dots in white), a few reverse micelles (white dots in black), and various bilayer and inverted bilayer structures that create a large amount of interface (amphiphiles, at the border between white and black). Fig. 12(c) illustrates the asymptotic effect of the successfully evolved replicators on that equilibrium. Note that the replicators have changed the self-assembled structures by filling the watery phase with micelles and filling the oily phase with reverse micelles. Comparison of 12(b) and 12(c) dramatizes the effect of coupling replicators and self-assembled structures. The replicators have changed the amphiphilic structures to create much more interface, which they need to reproduce, thus evolving a coupling with the amphiphiles that produces a favorable local environment for themselves.

< Fig. 13 Here >

Figure 13 illustrates key statistics measured as the replicators evolve. Note that the population shows signs of making two significant adaptations, the first around time step 300, and the second around time step 800. The timing of the first adaptation (around time step 300) corresponds to the onset of a steady fall in the mean value of gene 1; the corresponding characteristic rise and fall in the variance in gene 1 and the rise of the replicator concentration to its first plateau. The timing of the second adaptation (around time step 800) is confirmed by the slight final bump in mean value of gene 1 in figure 13 (d), when the new oily-living subpopulations grow and raise the mean gene 2 value; this timing is corroborated by the corresponding rise of the replicator concentration to its second plateau and the variance dynamics in that gene (not shown).
Figure 14 illustrates the evolutionary sequence. The amphiphiles and replicators together are shown on the left, and the amphiphiles only are shown on the right. We clearly see two separate kinds of adaptations. The first involves colonizing the watery niche by creating interface via micelles. The second involves colonizing the oily niche by creating interface via reverse micelles. Comparison of equilibrium spin configurations with and without replicators shows the dramatic effects of co-evolution of replicators and amphiphiles.

< Fig. 14 Here >

Figure 15 illustrates evolutionary activity waves[37,38] indicating the evolution of significant adaptations. The activity statistics were computed by measuring persistence of genes in the population, attaching a counter to each gene, and propagating the counter unchanged to offspring, except when a mutation occurs, when the counter is set to zero. Activity is plotted by forming the distribution of persistence counter values in the population, and observing the evolution of that distribution as a function of time. Formation and propagation of a wave in such a plot indicates the introduction of an innovation that persists because of survival value conferred on replicators by a particular gene value. The large wave A originating at the origin corresponds to certain combinatorial molecules that originated in the few initial replicators that eventually seeded the first successful population. The first adaptation corresponds to activity wave starting near time step 200, when the population learns to survive in the watery phase by mutations in gene 1 that have the effect of creating micelles and thus increasing the amount of interface. Successive adaptations continue, and the most successful version of this adaptation originates around time step 300, when the population level of replicators shows a significant increase. The second adaptation corresponds to activity wave originating around time step 600, when a small subset of the population migrates into the oily phase and learns to create reverse micelles. This figure confirms that the main adaptive events are not mere selection from variability that is present in the initial population, but instead involves new innovations. The many waves in starting arount time step 800 correspond to a succession of different subpopulations that independently migrate into the oily phase and adapt to it.

< Fig. 15 Here >

5. Discussion and conclusion

We have shown how the physics of complex fluids can be coupled with that of physicochemical evolution and reaction-diffusion kinetics to yield physicochemical evolution, and to elucidate the way in which pattern formation in collective physical phases (self-assembly) can become the object of natural selection. We have explored this path of combinatorial evolution from simple Ising models of phase transitions as far as spin-lattice models of ternary amphiphile systems such as the Widom model and beyond, which are known to capture essential features of the complex phase diagram of real emulsion systems.

The character of this paper is clearly that of opening up a large new area of inquiry with a presentation of first results, rather than an exhaustive analysis of all intermediate models. We have not attempted to utilize analytical treatments of the equilibrium properties of spin-lattice models in the analysis of the evolutionary system. It would be useful to have results on the equilibrium properties of the non-reacting versions of the coupled models we have presented of complex combinatorial fluids. The replica method and renormalization techniques[39] have proved successful in analyzing the steady state properties of combinatorial replicator models with statistical ensembles of fitness values, and could be applied here.

The evolution of pattern forming systems has been studied in the context of reaction-diffusion systems by one of us in previous works, both experimentally and theoretically[16], and this work has also yielded analytical tools[17], which may be relevant to the current context. It is clear that microreactor technology, with laser induced fluorescence detection, can be used to create controlled gradients of temperature and concentrations with which the phase space screens utilized for rapid insight in this work can be realized experimentally. The control of non-equilibrium self-assembly of binary and ternary fluids has been the subject of much recent investigation[40], not least because of potential biotechnological applications. Epstein has shown that emulsions can also be used to modify the properties of reaction-diffusion systems[41]. However chemical examples of the mutual interaction of complex fluid phases with immersed reaction-diffusion kinetics appear to be novel outside existing living cells. Because such systems can capture the essential physics of this coupling, they provide an important stepping-stone in the transition between inanimate and animate matter.

The spin-lattice approach suffers a serious drawback when it comes to modeling the dynamics of collective effects. No inertial effects are included in the dynamics, and so although collective movement of mesoscale structures can be observed, we cannot hope to obtain the correct scaling of diffusion of these structures with increasing size, nor hydrodynamic interactions, nor the appropriate response to external forces. In separate work, we are pursuing dissipative particle dynamics[4243] to create inertial evolvable models conserving the numbers of molecules in the system, but these more explicit off-lattice simulations are more time-consuming and there are still open questions about valid physical parameterization in pushing the integration time step well beyond that of molecular dynamics through stochastic integration.

We have not yet explored the influence of a specific free energy or chemical building block resource on the self-organization in detail. Our modeling framework is equipped with a resource plane, with diffusing resources, which are consumed during replication. All the simulations reported in this paper, however, were performed with immediate replenishment of a uniformly dense resource population. The pattern formation ability of resource depletion has been powerfully illustrated in the context of reaction-diffusion systems by the self-replicating spots in the Scott-Gray model[19].

Various authors have constructed theoretical models of components of protocells. Closest in spirit to the current approach is that of Ikegami[44,45] who employed a more complex orientation-dependent model (*cf.* [46]) of a membrane-forming system to analyze the chemical kinetics in complex fluids supporting cell-like entities. The spectrum of models presented here, with amphiphilic structure modeled by spin configurations on a lattice, is simpler, and not biased towards membrane systems. It enables a sharper characterization of energy couplings and free energy flows that drive pattern formation. An evolving system coupled to a non-evolving one in two layers was also explored in a different and non-thermodynamic context (the game of life CA) by Taylor [47]. In connection with the origin of life, although our work employs a sequential combinatorial representation of genetic information, this is not critical for many of the conclusions, so that the current approach may also be applicable to compositional information models of protocells as introduced by Lancet[48].

An important property of the current model is that combinatorial molecules must act together in a concerted fashion if they are to induce a local phase change (which then impacts either directly on replication rate or indirectly via mobility changes or induced aggregation on template survival). This means that the central problem of natural selection with cooperative functionality is present in the coupling of combinatorial selection with phase control. In order to quantify this relationship, one could analyze the minimum density of combinatorial molecules required to induce the phase change for a given level of parameter influence. Note that it is well known that the outcome of cooperative selection is very strongly dependent on the mobilities of the molecules carrying the heritable information: and is supported only for an intermediate range of mobilities[17]. It is an open question, whether a system can co-evolve mobility and cooperative functionality to place itself in this window. It does appear that part of the functionality of a protocell is to provide a strong control of the mobility of genetic information.

Acknowledgements

The bulk of the work was supported by the EU in the PACE integrated project, EC-FP6-IST-FET-IP-002035. All authors acknowledge travel grants from the Santa Fe Institute in 2003, European Center for Living Technology 2004. Additional support for S.R. 2005 came from the LDRD-DR Protocell Assembly (PAs) project. The corresponding author wishes to thank Manfred Eigen for his continuing support, and his research group BioMIP, especially T. Maeke and S. McCaskill, for their advice and consideration.

References

[1] Eigen, M., McCaskill, J.S., Schuster P. 1989 The molecular quasispecies. Adv. Chem. Phys. **75**, 149-263.

[2] Johns, G.C. and Joyce, G.F. 2005 The promise and peril of continuous in vitro evolution. J. Mol. Evol. **61**:253.

[3] Fontana, W., Schnabl, W. and Schuster, P. 1989 Physical aspects of evolutionary adaptation and optimization. Phys. Rev. A **40**(6) 3301-3321.

[4] Schuster, P. 2006 Prediction of RNA secondary structures: from theory to models and real molecules. Rep. Prog. Phys. **69** 1419–1477.

[5] Boerlijst, M. and Hogeweg P. 1991 Spiral wave structure in pre-biotic evolutin –hypercycles stable against parasites. Physica D **48**, 17-28.

[6] Cronhjort., M.B. and Blomberg, C. (1997) Cluster compartmentalization may provide resistance to parasites for catalytic networks. Physica D **101**, 289-298.

[7] Böddeker, B. 1995 Physical model for coevolution and simulation on a hardware programmable processor. Dipl. Thesis. in Physics, Univ. Göttingen, Germany. Superviser McCaskill, J.S. FSU Jena.

[8] Nielsen, S., Lopez, C.F., Srinivas, G. and Klein, M.L. 2004 Coarse grain models and the computer simulation of soft materials. J. Phys. Condens. Matter **16** R481-R512.

[9] Coveney, P., Emerton A., and Boghosian, B. 1996 Simulation of self-reproducing micelles using lattice-

gas automation. J. Amer. Chem. Soc. **118**(4) 10719-10724.
[10] Mayer, B. and Rasmussen, S. 2000 Dynamics and simulation of self-reproducing micelles. International J. Mod. Physics C, **11**(4) 827-836.
[11] Gompper, M. and Schick. M. 1991 Lattice model of microemulsions. Phys. Rev. B. **41**(13) 9148-9162.
[12] Elliott, R. , Szleifer, I.and Schick, M. 2006 Phase Diagram of a Ternary Mixture of Cholesterol and Saturated and Unsaturated Lipids Calculated from a Microscopic Model. Phys. Rev. Lett. **96** 098101.
[13] Colgate, S. A., Rasmussen, S., Solem, J. C. Lackner K. 2003 An astrophysical basis for a universal origin of life. Adv. Complex Systems, DOI:10.1142/S0219525903001079.
[14] Rasmussen S.**,** Chen L.**,** Nilsson M.**,** Abe S. 2003 Bridging Nonliving and Living Matter. Artificial Life, **9**(3) 269-316(48).
[15] Rasmussen, S., Chen, L., Deamer, D., Krakauer, D., Packard, N. Stadler, P. and M. Bedau, M. 2004 Transitions between nonliving and living matter, Science **303** 963-965.
[16] McCaskill, J.S. 1997. Spatially resolved in vitro molecular ecology. Biophys. Chem. **66** 145-158.
[17] McCaskill, J.S., Füchslin, R.M., Altmeyer, S. 2001 The stochastic evolution of catalysts in spatially resolved molecular systems. Biol.Chem. **382** 1343-1363.
[18] Altmeyer, S., Füchslin, R.M. and McCaskill, J.S. 2004 Folding stabilizes the evolution of catalysts. Artificial Life **10**(1), 23-38.
[19] Pearson, J.E. 1993 Complex patterns in a simple system. Science **261** 189-192.
[20] Ising, E. 1925 Beitrag zur Theorie des Ferromagnetismus. *Zeitschr.. f. Physik* **31**, 253-258.
[21] Widom, B. 1986 Lattice model of microemulsions. J. Chem. Phys. **84**(12) 6943-6954.
[22] Hansen, A., Schick, M. and Stauffer, D. 1991 Generalized Widom model of amphiphilic systems. Phys. Rev. A 44 3686-3691.
[23] Larson, R.G. 1988-1992 J.Chem. Phys. **89** 1642; **91** 2479; **96** 7904.
[24] Schick, M. & Shih, W. 1986 Spin-1 model of a microemulsion. Phys. Rev. B **34**(3) 1797-1801.
[25] Metropolis, N., Rosenbluth, A. Rosenbluth, M., Teller A., Teller E. 1953 J. Chem. Phys. **21** 1087.
[26] Morawietz, D., Chowdhury, D., Vollmar, S. and Stauffer, D. 1992 Physica A **187** 126.
[27] Anderson, P.W. 1958 Phys. Rev. **109** 1492.
[28] Turing, A.M. 1952 Chemical morphogenesis. Philos. Trans. Roy. Soc. B **237** 37.
[29] von Kiedrowski, G. 1986 Angew. Chem. **93** 932.
[30] Eigen, M. 1971 Self-organization of matter and the evolution of biological macromolecules. Naturwissenschaften **58**, 465-523.
[31] Eigen, M. and P. Schuster 1979 The hypercycle – a principle of natural self-organization. Springer-Verlag, Berlin. A combined reprint of Naturwissenschaften **64**: 541-565(1977); **65**:7-41, 341-369 (1978).
[32] Szathmary, E. and Demeter, L. 1987 Group selection of early replicators and the origin f life. J. Theor. Biol. **128** 463-486.
[33] Biebricher, C.K., M.Eigen and W.C. Gardiner, Jr. Kinetics of RNA replication: plus-minus asymetry and double strand formation. Biochemistry **23** (1984) 3186-3194.
[34] Anderson, P.W. 1983 Suggested model for prebiotic evolution: The use of chaos. Proc. Natl. Acad. Sci. USA **80** 3386-3390.
[35] Blum, H.F. 1962 On the origin and evolution of living machines. Am. Sci. **49** 474-501.
[36] Luisi, P.L. and F.J. Varela "Self-replicating micelles – a chemical version of minimal autopoeitic systems." Origin Life Evol. Biosph. **19**: 633-644.
[37] Bedau, M.A. and Packard N.H. 1991. Measurement of Evolutionary Activity, Teleology, and Life. In C. Langton, C. Taylor , D. Farmer, and S. Rasmussen, eds., *Artificial Life II*, Santa Fe Institute Studies in the Sciences of Complexity, Vol. X, pp. 431-461. Redwood City: Addison-Wesley.
[38] M. A. Bedau, E. Snyder, C. T. Brown, and N. H. Packard. 1997. A Comparison of Evolutionary Activity in Artificial Evolving Systems and the Biosphere. In P. Husbands and I. Harvey, *Proceedings of the Fourth European Conference on Artificial Life, ECAL97*, pp. 125-134. Cambridge: MIT Press.
[39] McCaskill, J.S. 1984 A localization threshold for macromolecular quasispecies from continuously distributed replication rates. J. Chem. Phys. **80**(10) 5194-5202.
[40] Thorsen, T. , Roberts, R. W. , Arnold, F. H. & Quake, S. R. 2001 Dynamic pattern formation in a vesicle-generating microfluidic device. *Phys. Rev. Lett.* **86**, 4163–4166.

[41] Vanag, V.K. and Epstein, I.R. 2001 Pattern Formation in a Tunable Reaction-Diffusion Medium: The BZ Reaction in an Aerosol OT Microemulsion. Phys. Rev. Lett. **87**, 228301-1-4.
[42] A. Buchanan, G. Gazzola, M. A. Bedau. 2006. Evolutionary design of a model of self-assembling chemical structures. In N. Krasnogor, S. Gustafson, D. Pelta and J.L. Verdegay, eds., *Systems Self-Assembly: multidisciplinary snapshots*. Elsevier. In press.
[43] Bedau M. A. and Brown, C. T. 1999 Visualizing evolutionary activity of genotypes. Artif. Life **5** 17-35.
[44] Ono, N. and Ikegami, T. 2000 Selection of catalysts through cellular reproduction. in R. Standish et al., eds Proc. Int. Conf. Artificial Life (ALIFE8) Cambridge Mass, 57-64.
[45] Madina, D., Ono, N. and Ikegami, T. 2003 Cellular Evolution in a 3D Lattice Artificial Chemistry in the proceedings of ECAL 03 (eds. W.Banzhaf et al., Dortmund, Springer) pp.59-68.
[46] Dawson, K.A. and Kurtovic, Z. 1990 Lattice model of self-assembly. J. Chem. Phys. **92**(9) 5473-5485.
[47] Taylor, T. 2004 Niche Construction and the Evolution of Complexity. In Artificial Life IX (Proceedings of the Ninth International Conference on the Simulation and Synthesis of Living Systems), J. Pollack, M. Bedau, P. Husbands, T. Ikegami and R. Watson (eds.), MIT Press, (pp. 375-380).
[48] Serge, D., Ben-Eli, D., Lancet, D. 2000 Compositional genomes, Proc. Natl. Acad. Sci USA **97** 4112-7.

Figure Captions

Fig. 1 The basic structure of Widom model, here on a 2D hexagonal lattice, showing dimeric ternary (amphiphile, oil, water) system with a typical local configuration and its corresponding spin representation, and showing the system configuration excluded from the model.

Figure 2 Gradient portraits of the Widom model. Two thermodynamic parameters are varied spatially to illustrate the different phases accessible in the model. In particular, z_1 and z_2 are varied along the horizontal and vertical axes, respectively. Note that the z1 and z2 scales are reflected about the midpoint of the x and y-axes. For example, moving left to right along the x-axis, the z1 value decreases from 4 to -4 at the midpoint and then increases back to 4. In this four-fold symmetry, the minimum z1 and z2 values are at the center of each figure. This symmetry removes edge artifacts from the simulation and allows reproducibility to be assessed.

Figure 3. The mutual influence of the combinatorial templates and the multiphase system provides a dynamic environment for evolution. The local phase of the substrate system determines both the template mobility and the template replication rate, while the template genes locally influence the ternary system, which in an energetically neutral manner can exchange material with the environment and thus change composition. The template replication process drives the thermodynamics of the composed genetic-ternary system out of equilibrium, as each replication step is associated with an energy cost.

Figure 4: Influence of static combinatorial molecules on Ising and Widom phase diagrams. The right hand plot is the binary Ising (lower) or ternary Widom fluid (upper), with the hydrophobic and hydrophilic states shown in black and blue. The K+, K- (upper) or J (lower) values are perturbed uniformly between ± 1. The activity coefficients z1 and z2 are set to zero. A vertical thermal gradient kT is applied: Widom (0.05 to 1.45), Ising (0.05 to 1.15). the plots (a) and (b) show the free energy costs (a) ΔF Widom and (b) $\Delta E / kT$ Ising for local state (spin) changes in the respective models. Note that the color scale goes from -10 to +10 (black to green). A critical temperature can be discerned at *ca*. kT=0.7 in (d) or 1.0 in (c), and is more apparent in the Monte Carlo dynamics.

Figure 5: Phase behavior of mobile combinatorially molecules coupled to an Ising fluid. (a) the equilibrium spatial distribution of combinatorial molecules, with brightness proportional to the magnitude of their perturbation of the two component fluid coupling constant J. (b) the two component fluid (as in Figure 4d). A vertical temperature gradient is applied, here ranging from 0.2 in the center to 1.2 at the top and bottom of the plots, visualizing the phase diagram. A phase transition is observed at a temperature of *ca.* 1.0 between a homogeneous mixed fluid phase at high temperature and an ordered mesoscale phase, consisting of three mesoscale structures: regions of hydrocarbon, water and liquid crystal. The Monte Carlo simulations on a 256x256 square lattice were run until equilibrium, constant free energy, was attained.

Figure 6: Phase plots of coupled dynamics of combinatorial molecules and complex fluid in the Widom model. We show four views of a single snapshot of the steady state system initialized with a random spin and combinatorial molecule configurations. A vertical temperature gradient with kT =0.2 in the center and 1.2 at the top and bottom is applied, along with a horizontal gradient in the amphiphile activity coefficient (logarithmic) from -2 in the center to 0 at the periphery.(a) $\Delta F / kT$ for local change to ternary fluid (spin flip) on the scale (-20 to 20) (b) Energetic mobility barriers $\Delta F / kT$ for vertical interchange of two combinatorial molecule neighbors. The horizontal interchange plot is very similar. (c) The non random arrangement of combinatorial molecules in the low mobility regime is apparent in this plot of K^- (ranging along with K^+ between -0.5 (black) and +0.5 (green)). (d) The oil (black) and aqueous (blue) phases with amphiphiles at all interfaces.

Figure 7 : Free energy densities per site in the course of the Widom model Monte Carlo simulation of Fig. 6. Dotted line : total free energy density. Dashed line : free energy density induced by combinatorial molecules. Solid line: non-genetic free energy density of spin system.

Figure 8 : Evolution of combinatorial molecules which increase the interface energy between two fluid components to induce phase separation (in the Ising model). (a) Starting equilibrium configuration in

vertical temperature gradient (0.25 in center to 2.25 at extremes) (b) t=400. Overlay of two component fluid state in blue and free energy of local state changes in green, on scale (-20 kT to 20kT). Mutant combinatorial molecules are already inducing phase separation and phase bias in the low temperature region. (c) Steady state configuration at t=5000, showing combinatorial molecules in dark red and phase separation to predominantly oily phase at all but highest temperatures. (d) $\Delta F / kT$ for a local state change at t=5000. (e) Combinatorial J values at t=5000 on scale (-1,1). Note that absent combinatorial molecules give black values. (f) Corresponding plot for a neutral gene, (one which has no effect on amphiphilic spin dynamics, but which undergoes the same replication process as those that do).

Figure 9 : Time course of evolution of phase separation in combinatorial Ising model. (a) Mean local density of oily phase, starting at unbiased probability of 0.5 as function of time (MC updates per lattice site) (b) Total free energy per lattice site is that induced by combinatorial molecules, approaching steady state (c) RMS variation range of ΔJ values of population of combinatorial molecules (d) Corresponding results for a neutral gene. The population starts with uniform population of sequences, which have zero influence.

Figure 10 : Time course of evolution of combinatorial surfactant system in Ising case. (a) mean hydrocarbon density per lattice site (dark-gray), mean template density (mid-gray), density of hydrocarbon at sites with combinatorial molecules (light-gray) (b) free energies per lattice site: total (dark-gray), from combinatorial molecules (mid-gray), of induced complex phase in the absence of templates (light-gray) (c) RMS variation of population of Ising coupling constants J with unary encoding, starting with uniformly population with J=0 (d) corresponding neutral gene development (lower and upper RMS limits shown).

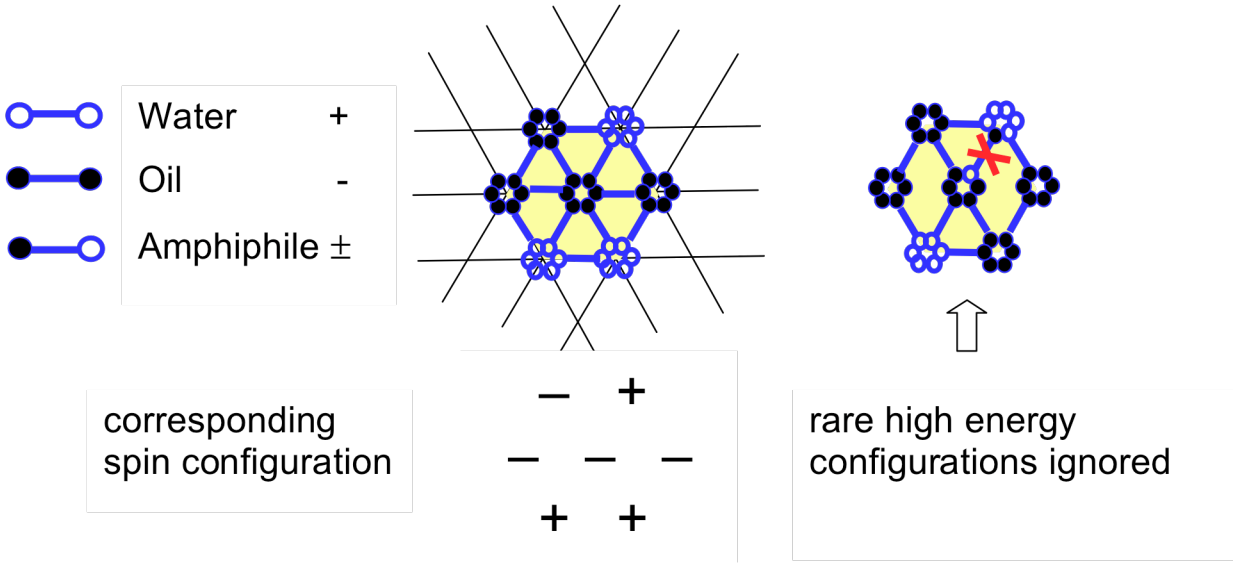
Figure 11 : Combinatorial Ising evolution of surfactant activity. Monte Carlo simulations as traced in Figure 10 with replication at the interface: with a vertical temperature gradient (0.5 in center and 1.0 at extremities). (a) Configuration after a short initial period t=500, showing phase separation: water (blue), oil(black), combinatorial molecules (red) (b) Free energy costs of state changes $\Delta F / kT$ for plot a (c) Representative example of evolutionary steady state (at t=45000) with almost complete conversion to fluctuating emulsion phase (d) $\Delta F / kT$ for evolved configuration in c (e) population of combinatorial molecules J values (scale -1 to 1 black to bright green), absent molecules black, showing lower negative values at low temperatures (f) neutral gene values on same scale. Simulation on 128x128 lattice with constant replication probability 0.5, loss probability 0.1, mutation rate 0.02, diffusion probability 0.1 and unary 30 bit landscape.

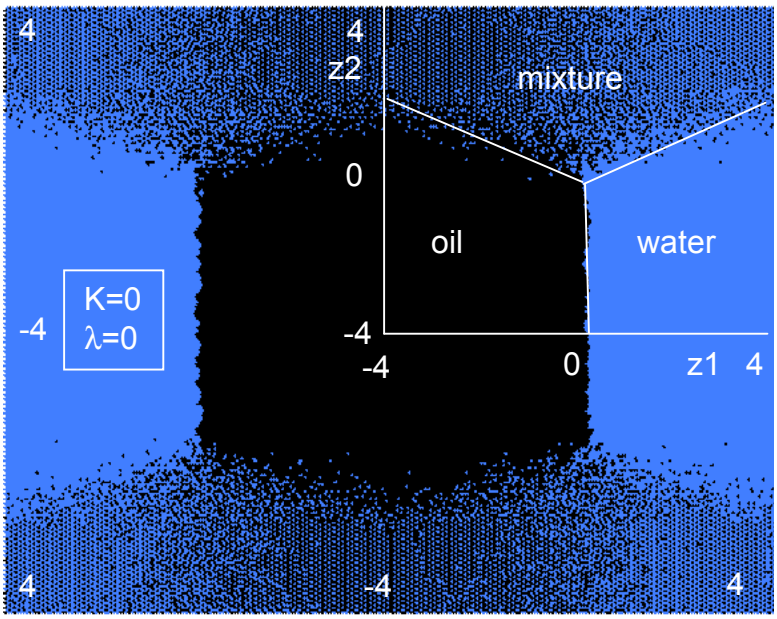
Figure 12. Three diagrams of the spin system (water, oil, amphiphile). (a) A short time after a random initial condition. (b) The equilibrium spin configuration (after a long time) when no replicators are present. (c) The steady state spin configuration after replicators that reproduce at the interface have colonized both the pure water and pure oil phases, as well as the interface. Simulations on a 150x150 hexagonal lattice. Replication, mutation, diffusion, and loss probabilities (0.5,0.1,1.0,0.0001) per MC update per lattice site.

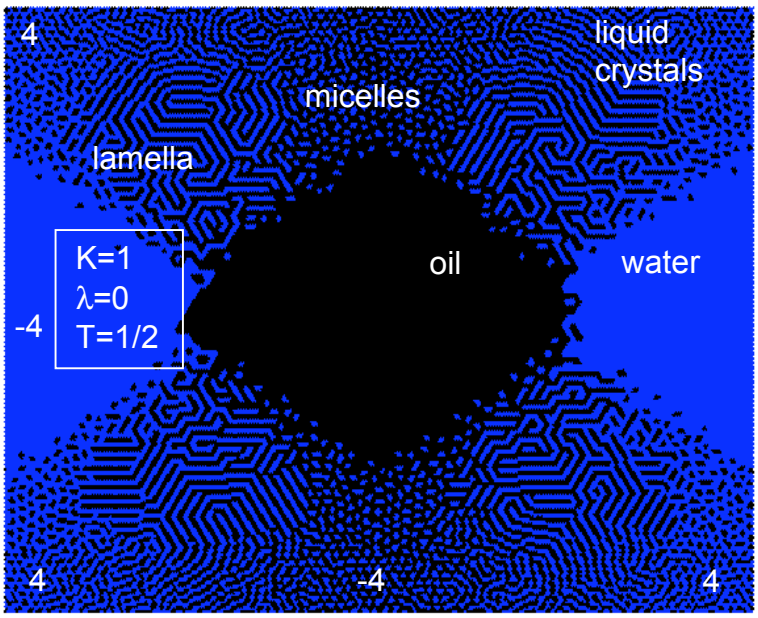
Figure 13. Time series of key statistics taken from the simulation. (a) The fraction of lattice sites that are hydrophilic. (b) The average density of combinatorial template molecules per lattice site. The population mean allele values at (c) gene 1, used in the aqueous phase and (d) gene 2, used in oily phase.

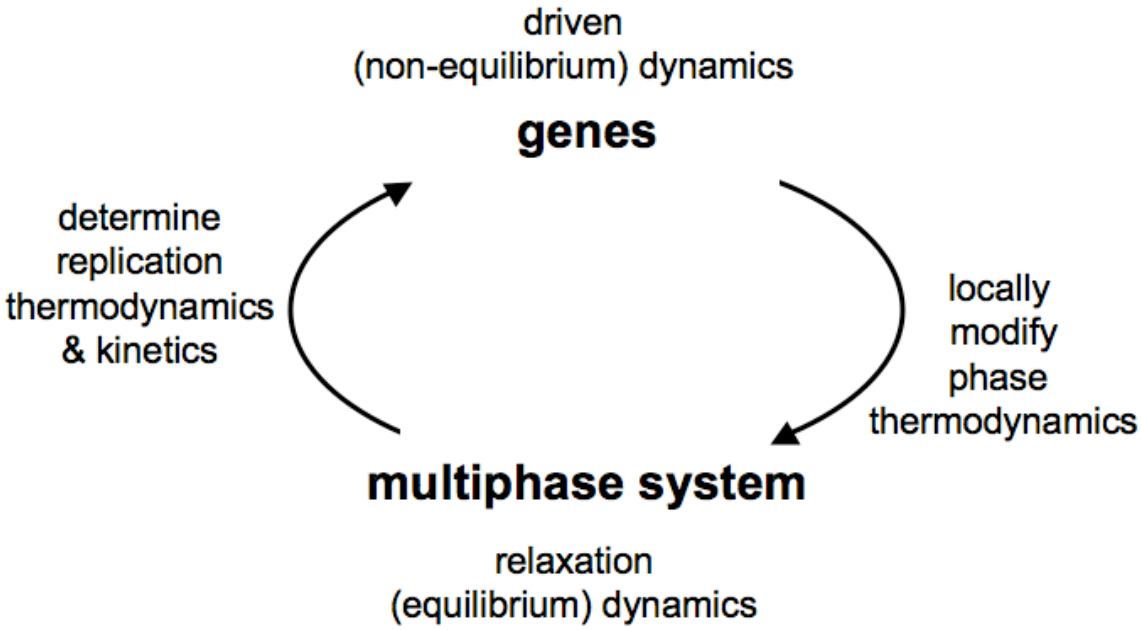
Figure 14 Evolutionary sequence of events associated with phase control accompanying Figures 12 and 13. (a) t=50 – a few replicators at the interface, oil/water relaxing to equilibrium (b) t=200 – replicators colonize the watery niche (c) t=600 – replicators have filled up the watery niche (d) t=700 three colonies get a foothold in the oily niche (two on left, one in middle) (e) t=800 – clearly see three colonies in oily niche (f) t=900 – replicators fill the oily niche.

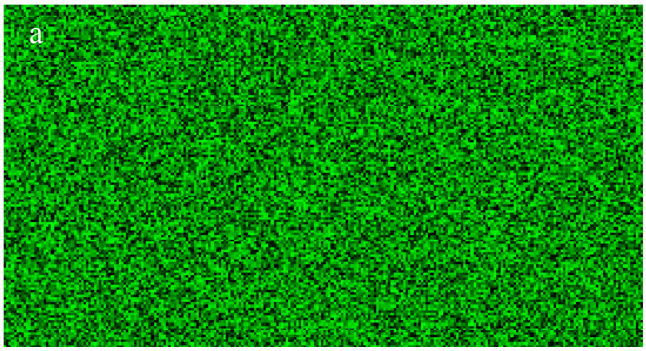
Figure 15. Dynamics in the evolutionary activity distribution as a function of time. The y-axis of the plot is activity measured as the persistence of genes surviving in the population, and the grey scale is proportional to the number of genes in the population having a given persistence. Timepoints of snapshots of the amphiphiles and replicators at three times Fig 14 b,c,e, show the correlation between the larger activity waves and major adaptation.

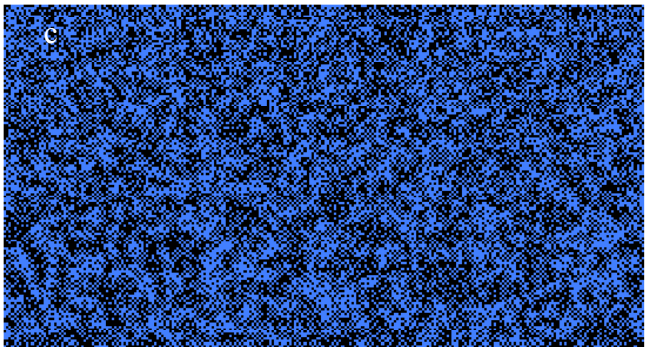

Water +
Oil -
Amphiphile ±
corresponding spin configuration
rare high energy configurations ignored

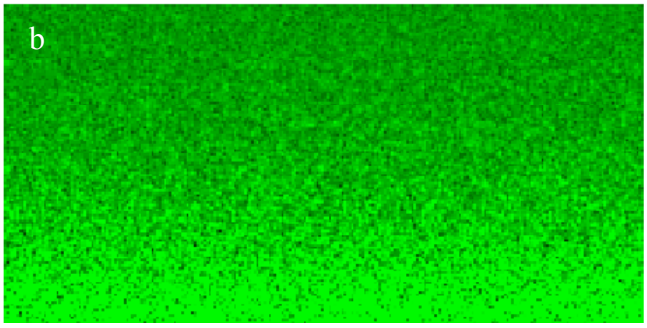

4
4
z2
mixture
0
oil
water
K=0
λ=0
-4
-4
-4
0
z1
4
4
-4
4

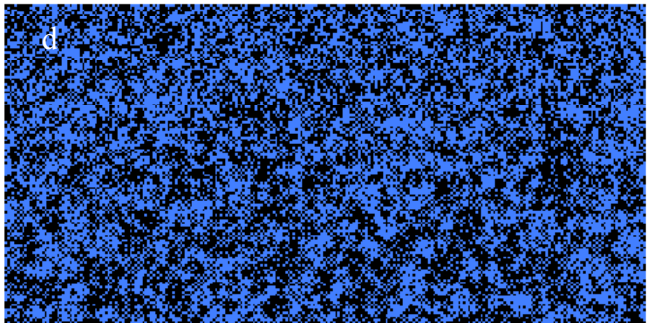

4
liquid
crystals
micelles
lamella
K=1
λ=0
T=1/2
oil
water
-4
4
-4
4

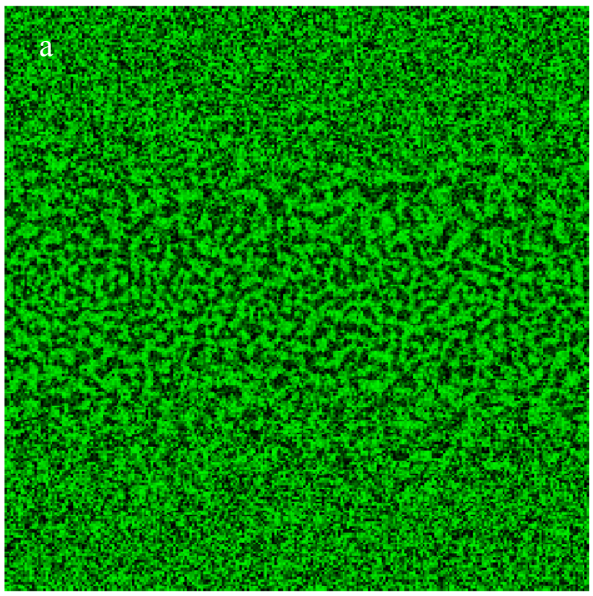

driven
(non-equilibrium) dynamics
genes
determine
replication
thermodynamics
& kinetics
locally
modify
phase
thermodynamics
multiphase system
relaxation
(equilibrium) dynamics

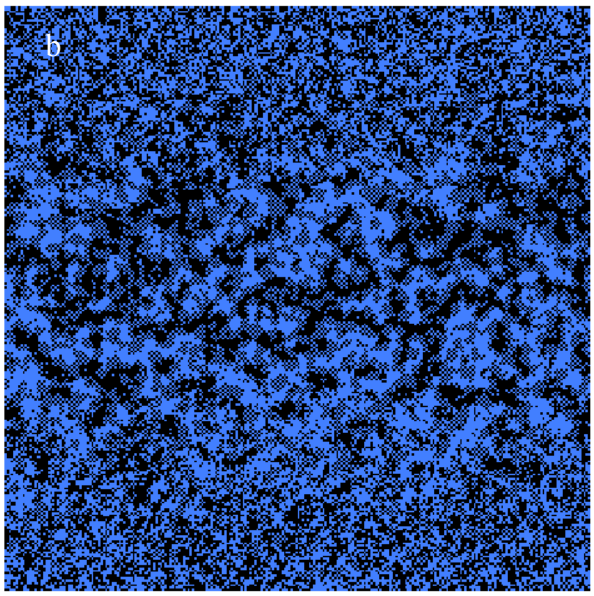

a

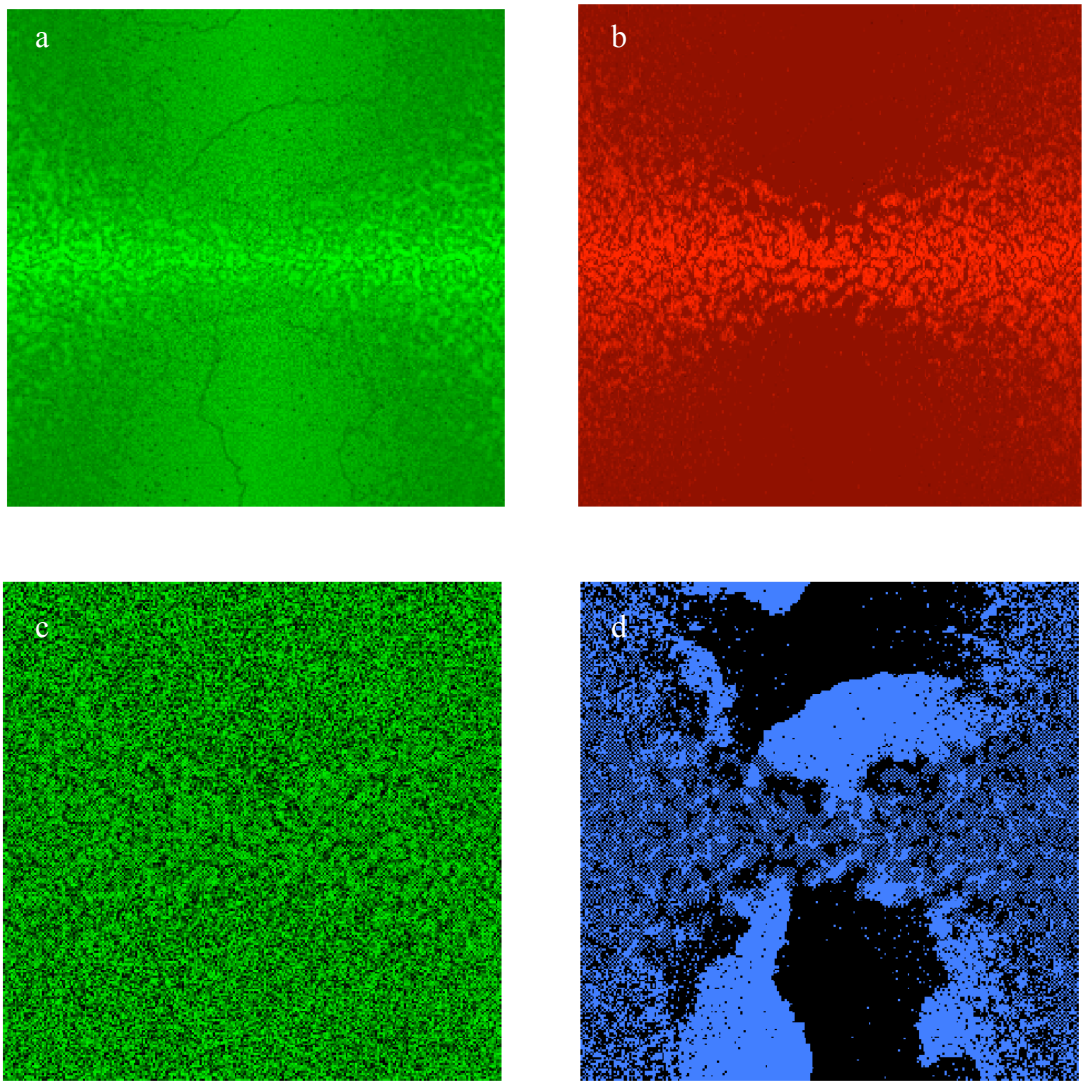

c

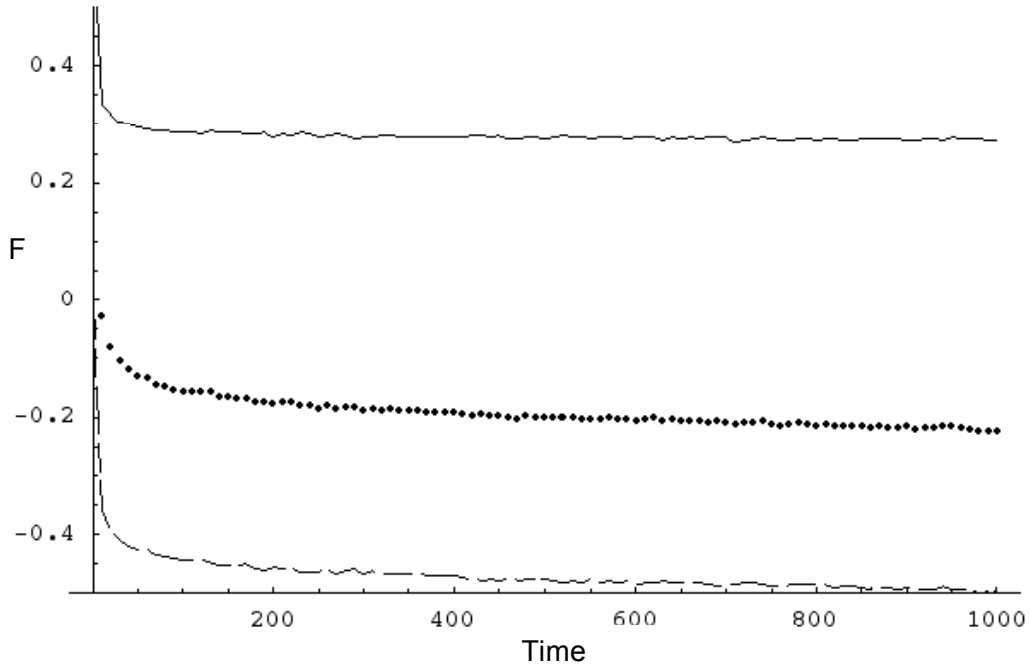

b

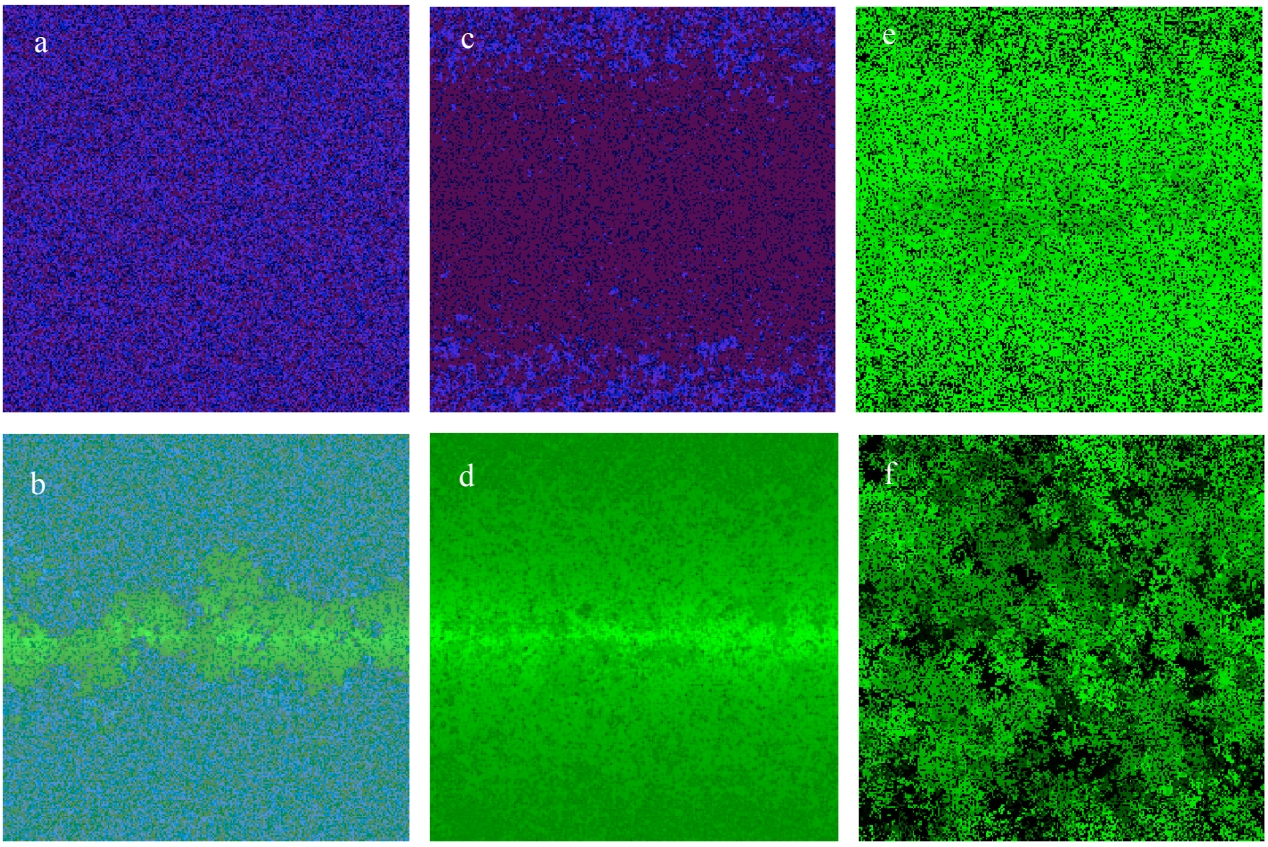

d

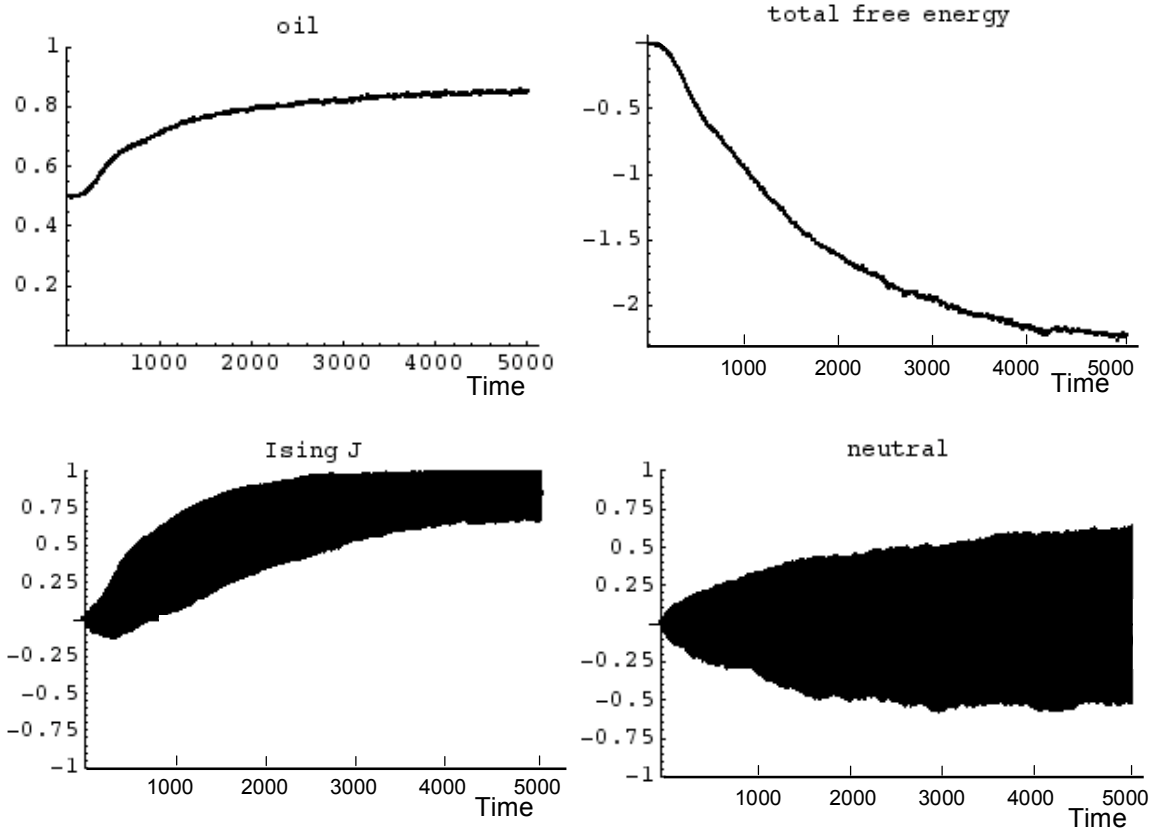

a

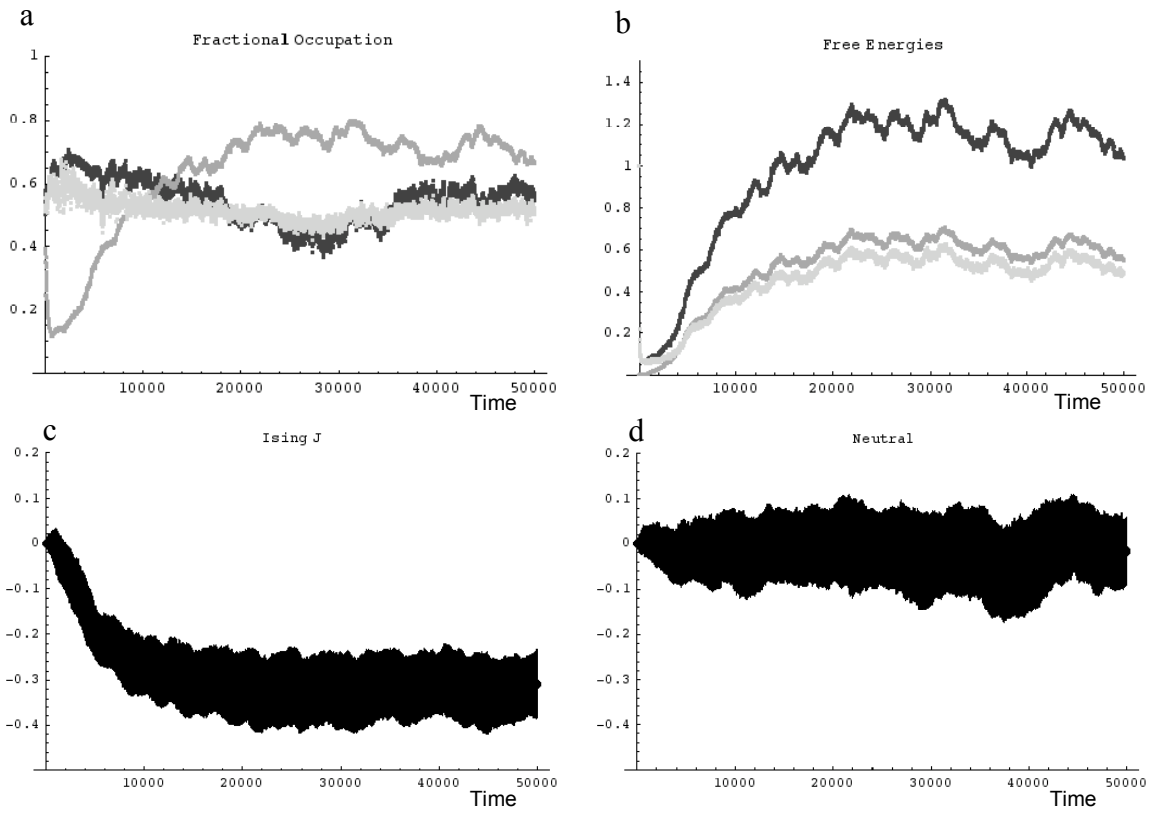

b

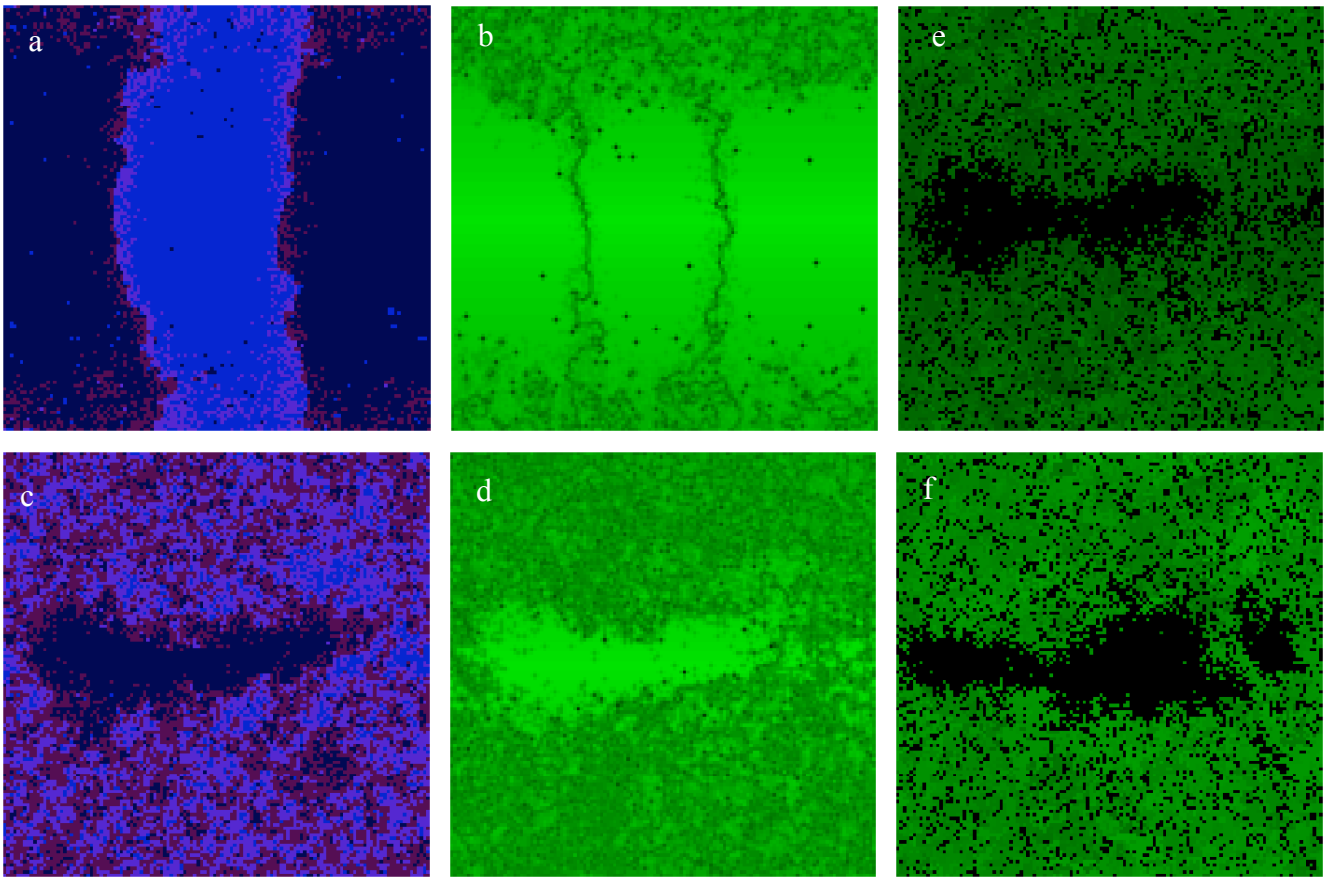

a
b
c
d


0.4
0.2
F
0
-0.2
-0.4
200
400
600
800
1000
Time

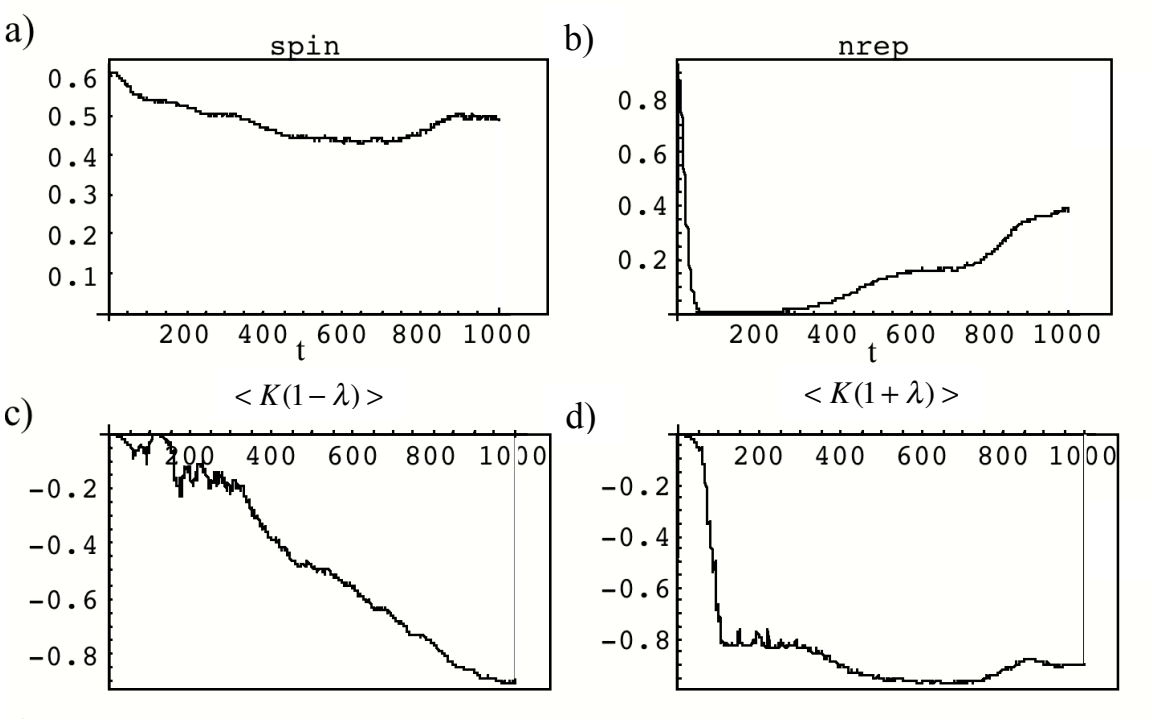

a
c
e
b
d
f

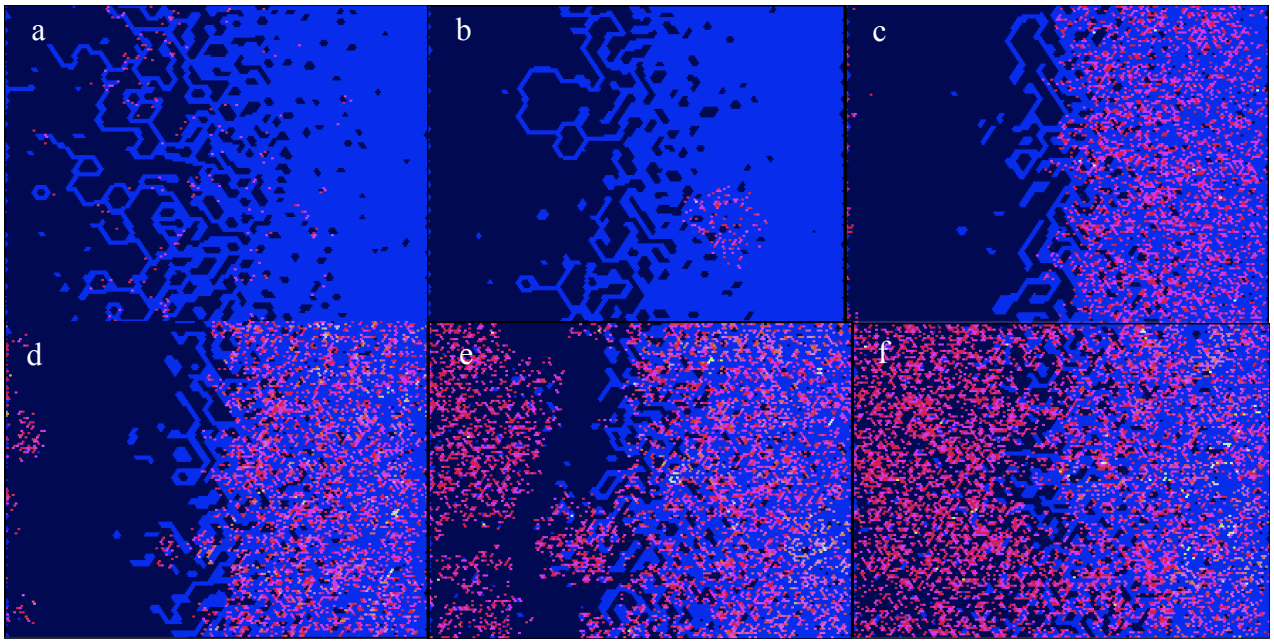

oil
1
0.8
0.6
0.4
0.2
1000
2000
3000
4000
5000
Time
total free energy
-0.5
-1
-1.5
-2
1000
2000
3000
4000
5000
Time
Ising J
1
0.75
0.5
0.25
-0.25
-0.5
-0.75
-1
1000
2000
3000
4000
5000
Time
neutral
1
0.75
0.5
0.25
-0.25
-0.5
-0.75
-1
1000
2000
3000
4000
5000
Time

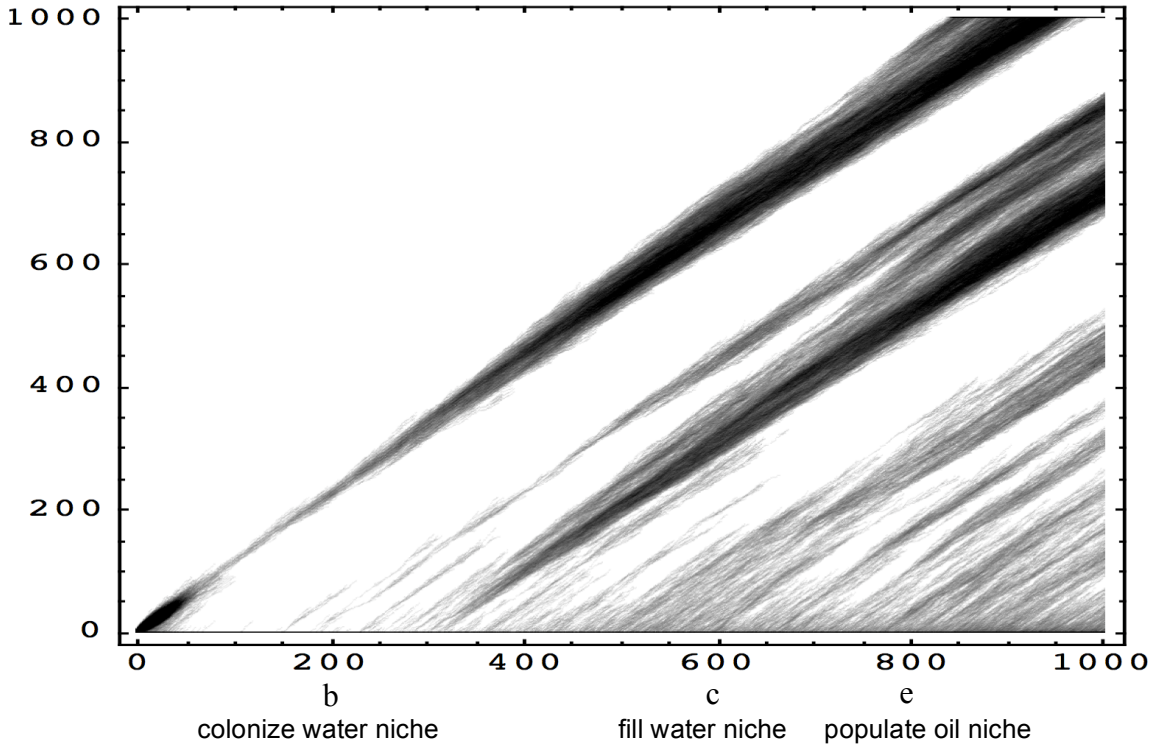

a
Fractional Occupation
1
0.8
0.6
0.4
0.2
10000
20000
30000
40000
50000
Time
b
Free Energies
1.4
1.2
1
0.8
0.6
0.4
0.2
10000
20000
30000
40000
50000
Time
c
Ising J
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
10000
20000
30000
40000
50000
Time
d
Neutral
0.2
0.1
0
-0.1
-0.2
-0.3
-0.4
10000
20000
30000
40000
50000
Time


a
b
e
c
d
f


a
b
c


a)
spin
0.6
0.5
0.4
0.3
0.2
0.1
200 400 600 800 1000
t
b)
nrep
0.8
0.6
0.4
0.2
200 400 600 800 1000
t
c)
< K(1 − λ) >
-0.2
-0.4
-0.6
-0.8
200 400 600 800 1000
d)
< K(1 + λ) >
-0.2
-0.4
-0.6
-0.8
200 400 600 800 1000


a
b
c
d
e
f


1000
800
600
400
200
0
0
200
400
600
800
1000
b
colonize water niche
c
fill water niche
e
populate oil niche